University of Alberta

PHYLOGENETICS OF ASPARAGALES (LILIOPSIDA) AS INFERRED FROM A LARGE PLASTID DNA SEQUENCE DATA SET

by

Marc Alexander McPherson ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of Master of Science
in Systematics and Evolution

Department of Biological Sciences

Edmonton, Alberta
Spring 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-82311-3

Canada

University of Alberta

Library Release Form

Name of Author: Marc Alexander McPherson

Title of Thesis: PHYLOGENETICS OF ASPARAGALES (LILIOPSIDA) AS INFERRED FROM A LARGE PLASTID DNA SEQUENCE DATA SET.

Degree: Master of Science

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

8708 78 Avenue
Edmonton, Alberta, Canada
T6C 0N6

Date: Jan 31/2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled PHYLOGENETICS OF ASPARAGALES (LILIOPSIDA) AS INFERRED FROM A LARGE PLASTID DNA SEQUENCE DATA SET by Marc Alexander McPherson in partial fulfillment of the requirements for the degree of Master of Biological Sciences (Systematics and Evolution).

Sean W. Graham (Supervisor)

Warren J. Gallin

Ruth A. Stockey

Ruth A. Stockey

Curtis Strobeck

Date: January 30, 2003

Abstract

To better understand deep phylogenetic relationships in Asparagales, the monocot order that includes the daffodils, irises, onions and orchids, I collected DNA sequence data from 24 of its ca. 29 families, spanning some twenty genes and eight disjunct regions of the plastid genome. The commelinoid monocots (gingers, grasses, palms and relatives) are inferred to be the sister group of Asparagales. The molecular data clarify the order's circumscription and permit robust determination of the branching history of most of its major lineages. For example, the astelioids (a predominantly Southern Hemisphere clade) and orchids define its two most basal evolutionary splits. Relationships within the "higher Asparagales" are mostly robustly resolved, but the Mediterranean endemic *Aphyllanthes monspeliensis* destabilizes the inference of some local relationships, explored here using Monte Carlo simulations. I also document the phylogenetic distribution and evolutionary significance of intron loss in a slowly evolving plastid locus (3-*rps*12) from two families of Asparagales.

Acknowledgments

I could not have accomplished this work without the support and love of my friends, parents, brother, wife and all of my children. Special thanks to Hardeep S. Rai for his continuing friendship and collaborative efforts on much of the work presented here. I would also like to thank the students of Dr. Curtis Strobeck's lab for their input on a project not presented here, especially Cory Davis for several afternoons of guidance and help. I would also like to thank Dr. John Bain for his patience, enthusiasm, inspiration and encouragement to pursue graduate studies in plant systematics.

The work presented here would not have been possible without the supervision of Dr. Sean Graham. I would like to acknowledge the helpful guidance of the members of my supervisory committee, Dr. Ruth Stockey and Dr. Curtis Strobeck. I thank Dorothy Fabijan (assistant curator of the herbarium), and Steven Williams (manager of the greenhouse facilities) for providing helpful advice and assistance with plant materials.

Many people contributed plant and / or DNA for my studies. Special thanks go to Chris Pires for some rare DNAs and an unpublished copy of his phylogeny of Themidaceae from his Ph.D. thesis.

I acknowledge support and funding from the Department of Biological Sciences Department for my graduate teaching assistantship and employment as a lab technician to help finance my final semester of studies. Special thanks go to Kim Christopher, Maggie Haag and the technical support staff for the undergraduate labs for their support and encouragement. I would also like to thank the molecular biology service unit (MBSU) staff (Bill Clark, Gary Ritzel, Pat Murray and Lisa Ostafichuk) for their kind help with techniques and equipment. I would also like to thank two volunteers in our lab, Winson Wong and Elvira Horne, for their hard work and dedication to the lab and to my project.

This work was funded in part by a research grant to Dr. Sean Graham from NSERC. The Challenge Grants in Biodiversity program and the Iris Society of America also provided financial support.

Dedicated to my wife Peggy and my children
Aliesha, Candace, Cody and Dustin

~~ TABLE OF CONTENTS ~~

CHAPTER 1. SYSTEMATICS OF THE MONOCOT ORDER ASPARAGALES: AN OVERVIEW 1
Bibliography 23

CHAPTER 2. PARALLEL LOSS OF AN INTRON IN SEVERAL CLOSELY RELATED MONOCOT FAMILIES (ASPARAGALES: ASPHODELACEAE AND HEMEROCALLIDACEAE)
Introduction 32
Materials and Methods 35
Results 44
Discussion 50
Conclusion 59
Bibliography 82

CHAPTER 3. PHYLOGENETIC RELATIONSHIPS AMONG THE MAJOR LINEAGES OF ASPARAGALES BASED ON A LARGE PLASTID DATA SET, WITH A SPECIAL FOCUS ON THE POSITION OF APHYLLANTHACEAE
Introduction 94
Materials and Methods 101
Results 114
Discussion 121
Bibliography 151

CHAPTER 4. CONCLUSION 162
Bibliography 167

List of Tables

Table 1.1. Comparison of three prominent monocot classification schemes 20

Table 2.1. Taxa examined for 3'-*rps*12 intron status in Asparagales 60

Table 2.2. Likelihood ratio test 65

Appendix. Intron status in groups outside of the Asparagales 76

Table 3.1. GenBank accession numbers and vouchers for exemplar monocot taxa 129

Table 3.2. Primers designed for this study 137

Table 3.3. Likelihood ratio test 138

Table 3.4 Estimated error rates for maximum parsimony inference given specified tree topologies 140

List of Figures

Figure 2.1. Overview of the plastid 3'-*rps*12 and *rps*7 operon 66

Figure 2.2. Amplification products for a subset of taxa surveyed in Asparagales for 3'-*rps*12 intron presence/ absence 68

Figure 2.3. One of four most-parsimonious trees of Asparagales inferred from a combined analysis of several chloroplast regions (*atp*B, *rbc*L and *trn*L-F) 70

Figure 2.4. Flat-weighted parsimony reconstructions of the evolution of 3'-*rps*12 intron status in a subset of Asparagales 72

Figure 2.5. Parallel intron loss reconstructed using an uneven parsimony weighting scheme on a suboptimal tree 74

Figure 3.1. Single most-parsimonious tree inferred for Asparagales and related monocots based on a large chloroplast data set 141

Figure 3.2. Portions of the trees used for Monte Carlo simulations, represented as parsimony phylograms 143

Figure 3.3. Single best maximum likelihood tree inferred for Asparagales and three commelinoid outgroup taxa 145

Figure 3.4. Branch labels scored in bootstrap analyses presented in Fig. 3.5 147

Figure 3.5. Spectrum of bootstrap support for branches found in the most parsimonious tree, and a few additional branches that had > 50% support from at least one data partition 149

~~ CHAPTER ONE ~~

SYSTEMATICS OF THE MONOCOT ORDER ASPARAGALES: AN OVERVIEW

Significance of the study

Charles Darwin characterized evolution as "descent with modification" (Darwin 1859: Chapter 4). However, it is only relatively recently that we have been able to address the descent component of evolutionary biology with any degree of confidence (e.g., Hennig, 1950; Edwards and Cavalli-Sforza, 1964; Felsenstein, 1985). This is in large part due to the development of methods for generating large amounts of phylogenetic data (primarily DNA sequences), and the emergence of the computational algorithms and processing power necessary for analyzing these data in a statistical framework (e.g., Felsenstein, 2001). Relatively massive progress has been made in our understanding of phylogenetic relationships among and within the major branches of the "Tree of Life," but very substantial areas of systematic relationship remain ambiguous or unresolved (e.g., Doyle, 1998; Graham *et al.*, 2002). These uncertainties are problematic for any scientist wishing to use these advances to address questions of organismal evolution (genomic or phenotypic), to create botanical classification schemes that substantially reflect evolutionary history, or to characterize plant biodiversity in an evolutionary framework.

The angiosperms (flowering plants) are one of the dominant life forms on land. They are the primary photosynthetic organisms on the land in terms of biomass and species richness (e.g., Judd *et al.*, 2002), and are a major part of the Earth's ecology and form the backbone of human civilization and economy. My thesis addresses a major branch of angiosperm phylogeny, the asparagalean monocots (Liliopsida: Asparagales). This order of monocots is one of the most species-rich and diverse groups of angiosperms. This is largely a function of a single family in the order, the orchids (Orchidaceae), which have ca. 20,000 species. The order Asparagales thus contains a substantial portion of total monocot and angiosperm species diversity (ca. 65,000 species of monocots and 257,000 species of flowering plants have been recognized; species numbers from Judd *et al.*, 2002). As currently circumscribed (Angiosperm Phylogeny Group, APG, 1998) this cosmopolitan order includes many taxa that are economically important or aesthetically pleasing to humans, including the daffodils and relatives in the Amaryllidaceae (ca. 870 species), irises and relatives (Iridaceae; ca. 1,750 species), onions and allies (Alliaceae; ca. 645 species), and, of course, the orchids.

Our understanding of monocot systematic relationships has changed enormously in the past decade, fueled by two pioneering molecular studies that used the conserved chloroplast (plastid) gene *rbc*L (Chase *et al.*, 1993; Duvall *et al.*, 1993) to broadly address angiosperm and monocot phylogenetics. These studies greatly clarified the taxonomic circumscription of monocot taxa at the family level and above, provided new evidence for their phylogenetic relationship to each other, and demonstrated the existence of some groups not conceived of using other earlier classes of data. These DNA-sequence based investigations have been a tremendous catalyst for subsequent phylogenetic research on the flowering plants generally, and monocots specifically, at a variety of taxonomic levels, using *rbc*L, other genes and morphology. These studies recently culminated in the ordinal level classification of the flowering plants by APG (1998). Their classification system is based primarily on analyses of DNA sequence data from three genes (two plastid genes, *atp*B, *rbc*L, and the nuclear small subunit ribosomal DNA locus, or "18S rDNA"). This classification system conflicts with one of the most widely accepted schemes of monocot classification, that of Cronquist (1981, 1988). One of the most spectacular areas of disagreement concerns the order Asparagales (Table 1.1).

A brief history of the classification of Asparagales

The order Asparagales Bromhead (1838) was delimited in its modern conception based on morphological data (primarily of seeds) by Huber (1969, 1977), more fully characterized by Rolf Dahlgren and colleagues using additional morphological and biochemical characters (Dahlgren, 1980; Dahlgren and Clifford, 1982; Dahlgren and Rasmussen, 1983; Dahlgren *et al.*, 1985), and most recently substantiated using molecular data (e.g., APG, 1998). The latter treatment was modified from the work of Bremer *et al.* (1995, 1996, and 1998) and incorporates the results of several molecular studies. [Note that the APG classification was slightly revised for the monocots by Chase *et al.* (2000a); however, as the classification of Asparagales was not modified there, I refer to the APG (1998) system throughout this chapter, unless otherwise stated.]

I summarize here (Table 1.1) monocot classification according to three prominent recent classification schemes (Cronquist, 1981, 1988; Dahlgren *et al.*, 1985; APG, 1998) from the perspective of exemplar genera that I use in a large-scale phylogenetic analysis of Asparagales, presented in Chapter 3 of this thesis. In terms of overall composition and familial division, the circumscriptions of Asparagales in the classifications of Dahlgren *et al.* (1985) and APG (1998) are largely congruent. However, these taxonomic systems differ substantially from recent classifications of the monocots proposed by other influential plant taxonomists, including Cronquist (1981, 1988; Table 1.1), Takhtajan (1980, 1997) and Thorne (e.g., 1983, 1992).

One of the major areas of dissent concerns how Cronquist and these others dealt with the "petaloid monocots" (superorder Liliiflorae in Dahlgren *et al.*, 1985), and in particular the lily family, Liliaceae. Although highly discordant with the modern evolutionary perspective (APG, 1998; Table 1.1) in this regard, Cronquist's treatment (1981, 1988) is a natural progression from the work of earlier taxonomists (such as Krause, 1930; reviewed in Dahlgren and Clifford, 1982 and Kubitzki *et al.*, 1998a, b). In contrast, Huber (1969) and Dahlgren *et al.* (1985) used rather narrow family and ordinal concepts in the petaloid monocots, to avoid the creation of "heterogeneous assemblages" (Dahlgren and Clifford, 1982; p. 14). They subdivided these monocots into five orders (Table 1.1), the largest being the Asparagales and Liliales. They also placed many of the liliaceous monocots (Liliaceae *sensu* Cronquist) in the former order (e.g., Table 1.1).

Several morphological features (primarily of seed anatomy) readily identify members of Asparagales (e.g., Judd *et al.*, 2002). Their seeds have a characteristic outer epidermis that is usually either obliterated (in fleshy fruited taxa), or present with a phytomelan crust (in dry-fruited taxa). They also have a completely collapsed inner (tegminal) seed coat. In contrast, seed coats of member of Liliales lack phytomelan, possess a well-developed outer epidermis, and have cellularity in the inner seed coat. Tepals typically do not have spots on them in Asparagales, but are quite often so in Liliales. Not all of these characters are completely consistent within Asparagales, nor exclusive to the order. For example, seeds of Orchidaceae and some other Asparagales (Dahlgren *et al.* 1985: p. 93) lack phytomelan, and spotted tepals are not unknown within Asparagales (in some orchids, for example). There have therefore either been several evolutionary convergences or reversals (homoplasy) in major characters that we now use to define Asparagales – or these characters may act as synapomorphies for only subsets of Asparagales.

Cronquist felt that the characters that distinguish the orders Asparagales and Liliales in the scheme of Dahlgren *et al.* (1985) scheme were rather vague and indefinite (Cronquist, 1988; p. 493). These characters include (for Asparagales) "…nonspotted tepals, nectaries in the septa of the ovary (instead of at the base of tepals or stamens), and sometimes anomalous secondary growth (vs. lack of secondary growth)" (Judd *et al.*, 2002: p. 254). He was apparently frustrated with Dahlgren *et* *al.*'s (1985) concept of these two orders, because of the lack of clear macromorphogical markers of ordinal status. His response was to lump many of these monocots together into a single family, Liliaceae, that incorporated the majority of the families now recognized in Asparagales (*sensu* APG, 1998), and which also included numerous taxa now recognized as belonging to other orders (e.g., Table 1.1). Cronquist did recognize that his Liliaceae was a "rather inchoate family," but he could not find a reasonable way of subdividing it. His morphological concept of Liliaceae included taxa that possess conspicuous, perfect flowers, typically with six stamens, with the ovary usually superior, and with raphides present in a subset of cells. In his view, members of Liliaceae are also "...geophytes with soft annual (rarely perennial) leaves, the stems herbaceous or nearly so and usually without secondary growth" (Cronquist 1988: p. 494).

Unfortunately, this rather loose description fits many families now recognized as belonging to Asparagales and other orders (Dahlgren and Clifford, 1982; Dahlgren *et al.*, 1985). Arthur Cronquist was one of the most influential plant taxonomists, and his system of plant classification continues to have a pervasive influence on much of the recent botanical, floristic and ecological literature. For example, the ongoing series of the Flora of North America (Oxford University Press) arranges the monocots according to his taxonomic treatment. However, his classification system for the monocots is clearly rejected and superceded by modern phylogenetic evidence (Dahlgren *et al.*, 1985, APG, 1998 – see Table 1.1; Chapters 2 and 3). Awareness of this fact has been rather slow in coming, outside the field of monocot systematics.

Recent studies on Asparagales have focussed on three primary areas: (1) Clarification of higher-order relationships, including confirmation of which taxa belong in the order; (2) Clarification of the taxonomic circumscription of individual families, and; (3) Reconstructions of character evolution, including determinations of the morphological characters (synapomorphies) that support major phylogenetic subdivisions within the order. I briefly review the recent literature concerning all three areas. The primary focus of my thesis was on the first area (higher-order relationships in Asparagales; Chapters 2, 3); but Chapter 2 also has implications for the taxonomic circumscription of the day lily family, Hemerocallidaceae, and its close relatives. These goals are briefly outlined at the end of this chapter.

Phylogenetic relationships in Asparagales and identification of its sister group

Recent molecular studies by Chase *et al.* (1995a; their data set supplemented in Rudall *et al.*, 1997) and Fay *et al.* (2000) provide the largest and most comprehensive

DNA sequence data sets analyzed to date for Asparagales. Chase *et al.* (1995a) performed an analysis of *rbc*L data from 83 species of Asparagales (and 89 other monocots) whereas Fay *et al.* (2000) analyzed sequence data for three plastid regions (*atp*B, *rbc*L, and *trn*L to *trn*F, including associated noncoding regions) for 64 species in Asparagales, and 14 outgroup taxa. Although large in terms of many molecular studies, these studies obviously include only a small fraction of the total number of species in the order. Nonetheless, their exemplar taxa were carefully chosen to represent the taxonomic breadth of Asparagales and relatives as currently understood in higher-level systematics, an approach that I necessarily follow in this thesis (Chapters 2, 3).

Tree support in the single gene studies of Chase *et al.* (1995a) and Rudall *et al.* (1997) was generally weak. Many internal nodes were supported by decay values (Bremer, 1988) of "d2" or less (Chase *et al.*, 1995a), or jackknife and bootstrap values < 50% (Rudall *et al.*, 1997). In a major departure from the classification schemes of Huber (1969) and Dahlgren *et al.* (1985), a clade in the molecular analyses that otherwise corresponded to their conception of Asparagales included Orchidaceae and Iridaceae, but with only weak to moderate support. However, Asparagales as a whole had very little support (decay value "d1"), and its position in the monocots was largely unresolved on the strict consensus tree of the multiple most-parsimonious trees. The study of Chase *et al.* (1995a) also sparked numerous taxonomic realignments at the family level, and indicated the existence of a major clade of "higher Asparagales" defined by successive microsporogenesis, that includes the onion and daffodil families (Alliaceae and Amaryllidaceae), the yucca family (Agavaceae), the asparagus family (Asparagaceae) and others.

The follow-up study of Fay *et al.* (2000) based on more chloroplast regions reinforced many of the higher-order relationships and familial circumscriptions found in Chase *et al.* (1995a), but generally with more convincing statistical support (in this case measured by bootstrap analysis; Felsenstein, 1985). The monophyly of the order as a whole, including Orchidaceae and Iridaceae, was strongly supported, as was its overall placement in the monocots. The Asparagales were depicted as the sister-group of the commelinoid monocots, a large and diverse group that includes the cattails, gingers, grasses, pineapples, palms and sedges. However, one aspect of the analysis performed in their study was problematic. In addition to doing phylogenetic analysis with flat weighting of their characters (weighting all characters equally), they also performed an analysis in which they successively weighted each character by its perceived homoplasy on a starting tree, with iterative tree searches performed in this manner until the homoplasy indices stabilized. The use of successive weighting is generally frowned upon by the modern phylogenetics community (Swofford *et al.*, 1996: pp. 500-503), because there is no guarantee that homoplasy indices of characters on a starting tree are closely related to their actual levels of homoplasy. Fay *et al.* (2000) performed bootstrap analyses using both standard flat-weighting and successive weighting; the former branch support values were only partly reported (and noted only in the text). They were often substantially lower than the values from the analysis using successively weighted characters, which were the ones reported in detail on a summary tree (their Fig. 2). I re-analyze their data in Chapter 2 and present the complete results of a bootstrap analysis that uses flat weighting of all characters. Felsenstein (1981, 1983) provides some theoretical justification for using this approach (at least for for slowly evolving characters) during parsimony analysis, even when characters do not actually change with equal probability. Empirical justification for this approach comes from the high degree of congruence that is typically observed among phylogenies inferred from different plastid genes using flat weights (e.g., Graham *et al.*, 1998).

In a larger study of monocots (Chase *et al.*, 2000a) based on a more or less complete data set for the three loci used by APG (1998; *atp*B, *rbc*L and 18S) for 126 taxa (including 43 species of Asparagales), Asparagales had poor bootstrap support (53-56%) and there was poor support (< 50%) for its putative sister-group relationship to the commelinoid monocots. The study of Fay *et al.* (2000) indicated that Orchidaceae may be the sister-group of the remaining members of Asparagales (< 50% bootstrap support), and suggested a new basal clade (also with < 50% bootstrap support) of Asparagales composed of five families: Asteliaceae, Blandfordiaceae, Boryaceae, Hypoxidaceae, and Lanariaceae. They referred to the latter clade as the "astelioid group." This group is discussed in more detail here in Chapter 3.

The inclusion of Iridaceae in Asparagales (somewhere close to the higher Asparagales) was strongly supported in the study of Fay *et al.* (2000), but there was ambiguity in its precise phylogenetic position. Other recent studies also support the inclusion of Iridaceae in Asparagales, including the morphological analysis of Rudall and Cutler (1995) and several molecular studies (Chase *et al.*, 1995a,b, 2000a; Fay *et al.*, 2000; Rudall *et al.*, 1997; Reeves *et al.*, 2001). These also depict Iridaceae as part of a "lower asparagoid" grade. The inclusion of Iridaceae in the Asparagales is consistent with the presence of septal nectaries in some members of the family, one of the key characters that Dahlgren *et al.* (1985: p. 131) used to distinguish Asparagales from Liliales. However, the precise phylogenetic relationship of Iridaceae to the other families in the "lower Asparagales" remains substantially unresolved (Goldblatt, 1990; APG, 1998; Rudall *et al.*, 1997; Chase *et al.*, 2000a; Fay *et al.*, 2000).

While the broad studies of Asparagales phylogeny by Chase *et al.* (1995a) and Fay *et al.* (2000) are generally highly congruent with the classification scheme of Dahlgren *et al.* (1985), a subset of families in Asparagales *sensu* Dahlgren *et al.* (1985) were found to be nonmonophyletic in the molecular studies. Several of these families (*sensu* Dahlgren *et al.*) were polyphyletic, including Alliaceae and Anthericaceae. These and other families have since been re-evaluated in additional studies that have led to substantially altered circumscriptions of some families (see below), and the erection of several new families (Anemarrhenaceae, Behniaceae, Boryaceae and Xeronemataceae; Chase *et al.*, 1996, 1997, 2000b; Conran *et al.*, 1997). The latter four families are small (monotypic or monogeneric), morphologically distinct, and represent distinct lineages in molecular analyses. The monogeneric family Xeronemataceae, for example, is the sister group of a clade consisting of the "higher asparagoids" plus three closely related families (Xanthorrhoeaceae, Hemerocallidaceae, and Asphodelaceae) (Chase *et al.*, 1995a, 2000a, b; Rudall *et al.*, 1997; Fay *et al.*, 2000).

Several families recognized by Dahlgren *et al.* (1985) are synonymized with others in Asparagales *sensu* APG (Table 1.1, footnote 5). These are Dracaenaceae, Eriospermaceae, Nolinaceae and Ruscaceae (accepted as synonyms of Convallariaceae); Funkiaceae (accepted as a synonym of Agavaceae); and Phormiaceae (accepted as a synonym of Hemerocallidaceae). These expanded families correspond to clades of closely related taxa in the *rbc*L-based study of Chase *et al.* (1995a).

Dasypogonaceae, Calectasiaceae and Hanguanaceae *sensu* Dahlgren *et al.* (1985) have now been largely removed from Asparagales to the commelinoid monocots (APG, 1998; see Table 1.1), with several genera from Dasypogonaceae *sensu* Dahlgren *et al.* (1985) retained in Asparagales as part of Laxmanniaceae. Studies examining cell walls for the presence of ester-linked ferulic acid, a compound that fluoresces under ultraviolet light, have shown that all commelinoids examined to date possess these compounds. This biochemical character is not known outside the commelinoid monocots and is apparently an unreversed synapomorphy for this large clade (Harris and Hartley, 1980; Rudall and Caddick, 1994; Harris, 2000). Members of Calectasiaceae, Dasypogonaceae and likely also Hanguanaceae (all *sensu* APG) also have these fluorescent compounds in their cell walls (Harris and Hartley, 1980;

Rudall and Caddick, 1994; Harris, 2000), consistent with their being commelinoid monocots and not members of Asparagales; the genera of Dasypogonaceae retained in Asparagales *sensu* APG (*i.e.*, *Acanthocarpus* Lehm., *Chamaexeros* Benth., *Lomandra* Labill. and *Romnalda* P. Stevens) all lack UV-fluorescent cell walls.

The Anthericaceae *sensu* Dahlgren *et al.* (1985) were found to be grossly polyphyletic in analyses of *rbc*L sequence data (Chase *et al.*, 1995a,b, 1996; Rudall *et al.*, 1997) and morphological analysis (Rudall and Cutler, 1995). This family was subsequently reduced to around eight genera (Conran, 1998), including *Anthericum* L. and *Chlorophytum* Ker Gawl. (spider plant), with the remaining genera moved into other existing or new families. For example, Boryaceae were erected from two genera of the tribe Boryeae of Anthericaceae *sensu* Keighery (1984), validated by Chase *et al.* (1997, and see Chase *et al.*, 1996). A redefined Laxmanniaceae (Chase *et al.*, 1996) now contains around fifteen genera previously placed by Dahlgren *et al.* (1985) in Asteliaceae, Anthericaceae, Dasypogonaceae (specifically *Lomandra* and relatives) and Luzuriagaceae (note that Luzuriagaceae is now recognized as a member of Liliales; APG, 1998).

The monogeneric day lily family Hemerocallidaceae (*sensu* Dahlgren *et al.* 1985) has been expanded to encompass eight genera of Johnsonieae [another tribe in

Anthericaceae *sensu* Keighery (1984) and Dahlgren *et al.* (1985)], and ten genera from Phormiaceae (see Chase *et al.*, 1996; APG, 1998; note that Phormiaceae is now recognized as a synonym of Hemerocallidaceae). The monophyly of the re-circumscribed Hemerocallidaceae (*sensu* APG, 1998) has been relatively well supported in several studies (Rudall *et al.*, 1997; Chase *et al.*, 2000a; Fay *et al.*, 2000). A clade consisting of Hemerocallidaceae, Asphodelaceae, and Xanthorrhoeaceae has also been found in several studies (Chase *et al.*, 1995a; Rudall *et al.*, 1997; Fay *et al.*, 2000). These relationships are addressed in more detail in Chapter 2 from the perspective of a large structural mutation that I encountered during my DNA-sequence based survey of the order.

The molecular results of Chase *et al.* (1995a) and Fay and Chase (1996) also indicated that the onion family, Alliaceae, is polyphyletic, one of the more surprising conclusions to come from molecular studies of Asparagales (e.g., Rahn, 1998). This family has since been substantially modified from its circumscription in Dahlgren *et al.* (1985) and elsewhere. One group of Alliaceae *sensu* Dahlgren *et al.* (1985), based solely on the genus *Agapanthus*, has been resurrected as Agapanthaceae (APG, 1998; Kubitzki, 1998c; and see Fay and Chase, 1996). Agapanthaceae appear to be closely related to Alliaceae and Amaryllidaceae, with the precise arrangement of these three families uncertain in the large study of Fay *et al.* (2000). A second group, centred on the onion-like *Brodiaea* Sm. complex, has been resurrected as the family Themidaceae. Themidaceae have been re-investigated using several plastid regions (Pires *et al.*, 2001; Pires and Sytsma, 2002). The family is strongly supported in these studies by bootstrap analyses (Felsenstein, 1985) as monophyletic, and may be closely related to Hyacinthaceae.

However, the relationship of Themidaceae and Hyacinthaceae to each other and the other "higher Asparagales" is uncertain (cf. Chase *et al.*, 2000a; Fay *et al.*, 2000; Pires *et al.*, 2001; Pires and Sytsma, 2002). This ambiguity may derive from uncertainty in the phylogenetic position of the monotypic family Aphyllanthaceae, which was excluded from some analyses (Pires *et al.*, 2001, Pires and Sytsma, 2002), supported as the sister group of Hyacinthaceae (Chase *et al.*, 1995a; Fay *et al.*, 2000 using successively weighted parsimony), or weakly supported as the sister group of Asparagaceae (Fay *et al.*, 2000, using equal weighted parsimony). The Mediterranean endemic *Aphyllanthes monspeliensis* L. appears to be on a long branch (*sensu* Felsenstein, 1978; Hendy and Penny, 1985) in phylogenetic trees of Asparagales, which may explain the uncertain inter-relationships of families in this local group of "higher Asparagales" (Fay *et al.*, 2000). This possibility is addressed in detail in Chapter 3.

Overview of the thesis

The overall goals of this thesis are: (1) to examine the evolution and taxonomic significance of an unusual structural mutation in a subset of closely related families in the order (Chapter 2), and; (2) to contribute a more robust understanding of relationships within the order Asparagales than has previously been possible, by obtaining a very large exemplar-based plastid sequence data set for detailed phylogenetic analysis (Chapter 3).

Although recent molecular and morphological studies have greatly improved our understanding of the composition and higher-order phylogenetic relationships in Asparagales, many points of relationships remain elusive, not least the position of the order in monocot phylogeny as a whole. In the most comprehensive recent studies, Chase *et al.* (2000a) noted that some internal branches in Asparagales were short, and suggested that more nucleotide data might help improve the resolution for several relationships, including the putative basal position of both the Orchidaceae and the astelioids. Fay *et al.* (2000) also suggested that the addition of more data might aid in elucidating the inference of Asparagales phylogeny. I build on their work here by adding substantially more DNA sequence data for a representative collection of Asparagales. My general aim is not to refine knowledge of the circumscription of individual families *per se* (these are now fairly well understood, e.g., Kubitzki, 1998a), but to sample the broad taxonomic and phylogenetic diversity of the order (as determined in recent studies) using an exemplar taxon approach.

In Chapter 2, I discuss the circumscription and relationship of three closely related families in Asparagales based on a large structural mutation (an intron loss) that has an odd phylogenetic distribution. The highly conserved 3'-*rps*12 intron appears to have been lost in parallel among these closely related families. This study examines the evolutionary and taxonomic significance of this putative parallel loss. In Chapter 3, I perform an exemplar-based phylogenetic analysis of Asparagales and relatives using a greatly enlarged plastid DNA sequence data set derived using primer sets developed primarily by Graham and Olmstead (2000). This chapter also contains an examination of the phylogenetic position of *Aphyllanthes* L., including a detailed determination of its potential for misleading phylogenetic analysis in the "higher Asparagales." The final chapter briefly summarizes the results of the studies presented in Chapters 2 and 3, and provides some directions for future research on Asparagales.

Table 1.1. Classification of the monocot genera used in this thesis (Chapter 3), according to three widely used taxonomic schemes. Genera of Asparagales *sensu* APG (1998) are noted in bold [1].

Cronquist (1988) [2] (5 subclasses)	Dahlgren, Clifford & Yeo (1985) (10 superorders)	APG (1998) [3, 4] (10 orders; 11 families completely or partly unplaced)
ALISMATIDAE (4 orders)	ALISMATIFLORAE	Acorales
Alismatales (3 families)	Alismatales (5 families)	Acoraceae
Alismataceae	Alismataceae	*Acorus*
Sagittaria	*Sagittaria*	
Najadales (10 families)	Najadales (8 families)	Alismatales (14 families)
Scheuchzeriaceae	Scheuchzeriaceae	Araceae
Scheuchzeria	*Scheuchzeria*	*Spathiphyllum*
		Alismataceae
ARECIDAE (4 orders)	ARECIFLORAE	*Sagittaria*
Arales (3 families)	Arecales	Scheuchzeriaceae
Acoraceae	Arecaceae	*Scheuchzeria*
Acorus	*Roystonea*	
Araceae		Asparagales (29 families)
Spathiphyllum	ARIFLORAE	Agavaceae
Arecales	Arales (2 families)	***Yucca***
Arecaceae (Palmae)	Araceae	Alliaceae
Roystonea	*Acorus*	***Allium***
	Spathiphyllum	Amaryllidaceae
COMMELINIDAE (7 orders)		***Narcissus***
Cyperales (2 families)	BROMELIIFLORAE (6 orders)	Anthericaceae
Poaceae (Gramineae)	Bromeliales	***Chlorophytum***
Oryza	Bromeliaceae	Aphyllanthaceae
Triticum	*Ananas*	***Aphyllanthes***
Zea	Philydrales	Asparagaceae
Typhales (2 families)	Philydraceae	***Asparagus***
Typhaceae	*Philydrum*	Asphodelaceae
Typha	Typhales (2 families)	***Asphodelus***
	Typhaceae	Asteliaceae
LILIIDAE	*Typha*	***Astelia***
Liliales (15 families)	Velloziales	Blandfordiaceae
Agavaceae	Velloziaceae	***Blandfordia***
Phormium	*Talbotia*	Boryaceae
Xeronema		***Alania***
Yucca	COMMELINIFLORAE (4 orders)	Convallariaceae
Cyanastraceae	Poales (7 families)	***Maianthemum***
Cyanastrum	Poaceae	Hemerocallidaceae
Dioscoreaceae	*Oryza*	***Hemerocallis***
Dioscorea	*Triticum*	***Phormium***
Iridaceae	*Zea*	Hyacinthaceae
Iris		***Muscari***
Sisyrinchium	LILIIFLORAE (5 orders)	Hypoxidaceae
Liliaceae	Asparagales (30 families) [5]	***Curculigo***
Alania	Agavaceae	Iridaceae
Allium	***Yucca***	***Iris***
Aphyllanthes	Alliaceae	***Sisyrinchium***
Anticlea	***Allium***	
Asparagus	***Muilla***	

Cronquist (1988) [2]	Dahlgren, Clifford & Yeo (1985)	APG (1998) [3, 4]
LILIIDAE/ Liliales/ Liliaceae (contd.)	LILIIFLORAE (contd.)	Asparagales (contd.)
Asphodelus	Amaryllidaceae	Ixioliriaceae
Astelia	***Narcissus***	***Ixiolirion***
Blandfordia	Anthericaceae	Lanariaceae
Chlorophytum	***Alania***	***Lanaria***
Curculigo	***Chlorophytum***	Laxmanniaceae
Hemerocallis	Aphyllanthaceae	***Lomandra***
Ixiolirion	***Aphyllanthes***	Orchidaceae
Lilium	Asparagaceae	***Amerorchis***
Maianthemum	***Asparagus***	***Coelogyne***
Muilla	Asphodelaceae	***Cypripedium***
Muscari	***Asphodelus***	Tecophilaeaceae
Narcissus	Asteliaceae	***Cyanastrum***
Philydraceae	***Astelia***	Themidaceae
Philydrum	Blandfordiaceae	***Muilla***
Stemonaceae	***Blandfordia***	Xanthorrhoeaceae
Stemona	Convallariaceae	***Xanthorrhoea***
Velloziaceae	***Maianthemum***	Xeronemataceae
Talbotia	Cyanastraceae	***Xeronema***
Xanthorrhoeaceae	***Cyanastrum***	
Dasypogon	Dasypogonaceae	Dioscoreales (5 families)
Lomandra	*Dasypogon*	Burmanniaceae
Xanthorrhoea	***Lomandra***	*Burmannia*
Orchidales (4 families)	Hemerocallidaceae	Dioscoreaceae
Burmanniaceae	***Hemerocallis***	*Dioscorea*
Burmannia	Hyacinthaceae	
Orchidaceae	***Muscari***	Liliales (9 families)
Amerorchis	Hypoxidaceae	Liliaceae
Coelogyne	***Curculigo***	*Lilium*
Cypripedium	Ixioliriaceae	Melanthiaceae
	Ixiolirion	*Anticlea*
ZINGIBERIDAE	Phormiaceae	
Bromeliales	***Phormium***	Pandanales (4 families)
Bromeliaceae	***Xeronema***	Stemonaceae
Ananas	Tecophilaeaceae	*Stemona*
Zingiberales (8 families)	***Lanaria***	Velloziaceae
Musaceae	Xanthorrhoeaceae	*Talbotia*
Ensete	***Xanthorrhoea***	
	Burmanniales (3 families)	COMMELINOIDS [6]
NOT NOTED	Burmanniaceae	Unplaced at ordinal level (6
families)		
Lanaria	*Burmannia*	Bromeliaceae
	Dioscoreales (7 families)	*Ananas*
	Dioscoreaceae	Dasypogonaceae
	Dioscorea	*Dasypogon*
	Stemonaceae	
	Stemona	Arecales
		Arecaceae
		Roystonea

Cronquist (1988) [2]	Dahlgren, Clifford & Yeo (1985)	APG (1998) [3, 4]
	LILIIFLORAE (contd.)	COMMELINOIDS (contd.)
	Liliales (10 families, incl. orchids)	Commelinales (4 families)
	Iridaceae	Philydraceae
	Iris	*Philydrum*
	Sisyrinchium	
	Liliaceae	Poales (16 families)
	Lilium	Poaceae
	"Orchids" (rankless in Liliales; 3 families)	*Oryza*
	Cypripediaceae	*Triticum*
	Cypripedium	*Zea*
	Orchidaceae	Typhaceae
	Amerorchis	*Typha*
	Coleogyne	
	Melanthiales (2 families)	Zingiberales (8 families)
	Melanthiaceae	Musaceae
	Anticlea	*Ensete*
	ZINGIBERIFLORAE	
	Zingiberales (8 families)	
	Musaceae	
	Ensete	

---

[1] Taxon authorities are listed in Brummitt (1992) and APG (1998). The number of orders and families are noted in parentheses in instances where exemplar species sampling was incomplete at these higher taxonomic levels.

[2] Generic placement follows Cronquist (1981, 1988) and Brummitt (1992).

[3] Generic placement follows Kubitzki (1998a,b), except for Asparagales, where we follow Fay *et al.* (2000). The families of Asparagales recognized by APG (1998) but not sampled here are Agapanthaceae, Anemarrhenaceae, Behniaceae, Doryanthaceae, Herreriaceae and Hesperocallidaceae.

[4] An updated version of this classification system ("APG II"; M.W. Chase, pers. comm.) permits optional recognition of (i) Amaryllidaceae as a separate family or as a synonym of Alliaceae (together with Agapanthaceae); (ii) Agavaceae, Aphyllanthaceae, Hyacinthaceae, Laxmanniaceae and Themidaceae as separate families (or as synonyms of Asparagaceae, together with Hesperocallidaceae and Ruscaceae; the latter not recognized in APG, 1998, but optionally recognized here); (iii) Asphodelaceae and Hemerocallidaceae as separate families (or as synonyms of Xanthorrhoeaceae).

[5] Families of Asparagales *sensu* Dahlgren *et al.* (1985) that are not listed under the order here are: Calectasiaceae and Hanguanaceae (both unplaced commelinoids in APG, 1998); Doryanthaceae and Herreriaceae (not sampled here); Dracaenaceae, Eriospermaceae, Nolinaceae and Ruscaceae (synonyms of Convallariaceae in APG, 1998); Funkiaceae (synonym of Agavaceae in APG, 1998); Luzuriagaceae and Philesiaceae (transferred to Liliales by APG, 1998); Phormiaceae (synonym of Hemerocallidaceae in APG, 1998).

[6] The "commelinoid monocots" in the APG system comprise four orders (Arecales, Commelinales, Poales and Zingiberales); they correspond approximately to Areciflorae + Bromeliiflorae + Commeliniflorae + Zingiberiflorae in the system of Dahlgren *et al.* (1985).

Bibliography

Angiosperm Phylogeny Group., 1998. An ordinal classification for the families of flowering plants. Ann. Missouri Bot. Gard. 85, 531–553.

Angiosperm Phylogeny Group (APGII)., unpublished. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants.

Bremer, K., Bremer, B., and Thulin, M., 1995, 1996 and 1998. Introduction to phylogeny and systematics of flowering plants, First – third editions, Compendium. Uppsala University, Uppsala.

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42, 795–803.

Bromhead, E. Ff., 1838. An attempt to ascertain the characters of the botanical alliances. Edinburgh New Philos. J. 25, 123-133.

Brummitt, R.K. (comp.)., 1992. Vascular Plant Families and Genera. A listing of the genera of vascular plants of the world according to their families, as recognised in the Kew Herbarium, with an analysis of relationships of the flowering plant families according to eight systems of classification. Royal Botanic Gardens, Kew.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedrén, M., Gaut, B. S., Jansen, R. K., Kim, K.-J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q.–Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, Jr. G. H., Graham, S. W., Barrett, S. C. H., Dayanandan, S., and Albert, V. A., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80, 528–580.

Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J.G., Cox, A. V., Eguiarte, L. E., Hartwell, J., Fay, M. F., Caddick, L. R., Cameron, K. M., and Hoot, S., 1995a. Molecular phylogenetics of Lilianae. In: Rudall, P. J., Cribb, P. J., Cutler, D. F., and Humphries, C. J., (Eds.), Monocotyledons: Systematics and Evolution. Royal Botanic Gardens, Kew, Vol. 1, pp. 109–137.

Chase, M. W., Stevenson, D. W., Wilkin, P., and Rudall, P. J., 1995b. Monocot systematics: a combined analysis. In: Rudall, P. J., Cribb, P. J., Cutler, D. F., and Humphries, C. J., (Eds.), Monocotyledons: Systematics and Evolution. Royal Botanic Gardens, Kew, Vol. 2, pp. 685–730.

Chase, M. W., Rudall, P. J., and Conran, J. G., 1996. New circumscriptions and a new family of asparagoid lilies: genera formerly included in Anthericaceae. Kew Bull. 51, 667–680.

Chase, M. W., Rudall, P. J., and Conran, J. G., 1997. Validation of the family name Boryaceae. Kew Bull. 52, 416.

Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., Sullivan, S., Joseph, J., Molvray, M., Kores, P.J., Givnish, T. J., Sytsma, K. J., and Pires, J. C., 2000a. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 3–16.

Chase, M. W., Rudall, P. J., Fay, M. F., and Stobart, K. L., 2000b. Xeronemataceae, a new family of asparagoid lilies from New Caledonia and New Zealand. Kew Bull. 55, 865–870.

Conran, J. G., 1998. Anthericaceae. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp. 114–121.

Conran, J. G., Chase, M. W., and Rudall, P. J., 1997. Two new monocotyledon families: Anemarrhenaceae and Behniaceae (Lilianae: Asparagales). Kew Bull. 52, 995–999.

Cronquist, A., 1981. An integrated system of classification of flowering plants. Columbia University Press. N.Y. USA.

Cronquist, A., 1988. The evolution and classification of flowering plants, Second edition. The New York Botanical Garden, Bronx, N.Y. USA. pp. 451–517.

Dahlgren, R. M. T., 1980. A revised system of classification of the angiosperms. Bot. J. Linn. Soc. 80, 91–124.

Dahlgren, R. M. T., and Clifford, H. T., 1982. The monocotyledons: a comparative study. Springer-Verlag, Berlin.

Dahlgren, R. M. T., and Rasmussen, F. N., 1983. Monocotyledon evolution: characters and phylogenetic estimation. Evol. Biol. 16, 255–395.

Dahlgren, R. M. T., Clifford, H. T., and Yeo, P. F., 1985. The families of monocotyledons: structure, evolution, and taxonomy. Springer-Verlag, Berlin.

Darwin, C., 1859. The origin of species by means of natural selection or the preservation of favoured races in the struggle for life. Murray, London.

Doyle, J. A., 1998. Phylogeny of vascular plants. Ann. Rev. Ecol. Syst. 29, 567–599.

Duvall, M. R., Clegg, M. T., Chase, M. W., Clark, W. D., Kress, W. J., Hills, H. G., Eguiarte, L. E., Smith, J. F., Gaut, B. S., Zimmer, E. A., and Learn, G. H., 1993. Phylogenetic hypothesis for the monocotyledons constructed from *rbc*L sequence data. Ann. Missouri Bot. Gard. 80, 607–619.

Edwards, A. W. F, and Cavalli-Sforza, L. L., 1964. Reconstruction of evolutionary trees. In: Heywood, V. H., and McNeill, J. (Eds.), Phenetic and phylogenetic classification. Systematics Association Publ. No. 6, London, pp. 67-76.

Fay, M. F., and Chase, M. W., 1996. Resurrection of Themidaceae for the *Brodiaea* alliance, and recircumscription of Alliaceae, Amaryllidaceae and Agapanthoideae. Taxon 45, 441–451.

Fay, M. F., Rudall, P. J., Sullivan, S., Stobart, K. L., de Bruijn, A. Y., Reeves, G., Qamaruz-Zaman, F., Hong, W. -P., Joseph, J., Hahn, W. J., Conran, J. G., and Chase, M. W., 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 360–371.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Felsenstein, J., 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16, 183–196.

Felsenstein, J., 1983. Parsimony in systematics: biological and statistical issues. Ann. Rev. Ecol. Syst. 14, 313–333.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 2001. The troubled growth of statistical phylogenetics. Syst. Biol. 50, 465–467.

Fitch, W. M., 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Zool. 20, 406–416.

Goldblatt, P., 1990. Phylogeny and classification of Iridaceae. Ann. Missouri Bot. Gard. 77, 607–627.

Graham, S. W., Kohn, J. R., Morton, B. R., Eckenwalder, J. E., and Barrett, S. C. H., 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. Syst. Biol. 47, 545–567.

Graham, S. W., and Olmstead, R. G., 2000. Utility of 17 plastid genes for inferring the phylogeny of the basal angiosperms. Amer. J. Bot. 87, 1712–1730.

Graham S. W., Barrett, S. C. H., and Olmstead, R. G., 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. Mol. Biol. Evol. 19, 1769–1781.

Hendy, M. D., and Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Harris, P. J., 2000. Compositions of monocotyledon cell walls: implications for biosystematics. . In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 114–126.

Harris, P. J., and Hartley, R. D., 1980. Phenolic constituents of the cell walls of monocotyledons. Biochem. Syst. Ecol. 8, 153–160.

Hennig, W., 1950. Grundzüge einer Theorie der phylogenetischen Systematik. Deutscher zentralverlag, Berlin.

Huber, H., 1969. Die Samenmerkmale und Verwandtschaftsverhältnisse der Liliiflorae. Mitteilungen der botanischen Staatssammlung München 8, 219–538.

Huber, H., 1977. The treatment of the monocotyledons in an evolutionary system of classification. Plant Syst. Evol. (Suppl. 1), 285–298.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., Donoghue, M. J., 2002. Plant systematics: a phylogenetic approach, Second edition. Sinauer Associates, Sunderland, MA.

Keighery, G. J., 1984. The Johnsonieae (Liliaceae): biology and classification. Flora 175, 103–108.

Krause, K., 1930. Liliaceae. In Engler, A., (Ed.), Die natürlichen Pflanzenfamilien, Second edition, Volume 15a. W. Engelmann, Leipzig, pp. 227–386.

Kubitzki, K., 1998a. The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York.

Kubitzki, K., 1998b. The Families and Genera of Vascular Plants, Volume IV, Flowering plants. Monocotyledons: Alismatanae and Commelinanae (except Gramineae). Springer-Verlag, New York.

Kubitzki, K., 1998c. Agapanthaceae. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp. 58–60.

Kubitzki, K., Rudall, P. J., and Chase, M. W., 1998. Systematics and evolution. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp. 23–33.

Pires, J. C., Fay, M. F., Davis, W. S., Hufford, L., Rova, J., Chase, M. W., and Sytsma, K. J., 2001. Molecular and morphological phylogenetic analyses of Themidaceae (Asparagales). Kew Bull. 56, 601–626.

Pires, J. C., and Sytsma, K. J., 2002. A phylogenetic evaluation of a biosystematic framework: *Brodiaea* and related petaloid monocots (Themidaceae). Amer. J. Bot. 89, 1342–1359.

Rahn, K., 1998. Alliaceae. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp.70–78.

Reeves, G., Goldblatt, P., Rudall, P. J., Souza-Chies, T., Lejeune, T., Fay, M. F., Cox, A. V., and Chase, M. W., 2001. Molecular systematics of Iridaceae: evidence from four plastid DNA regions. Amer. J. Bot. 88, 2074–2087.

Rudall, P. J., and Caddick, L. R., 1994. Investigation of the presence of phenolic compounds in monocotyledonous cell walls, using UV fluorescence microscopy. Ann. Bot. 74, 483–491.

Rudall, P. J., and Cutler, D. F., 1995. Asparagales: a reappraisal. In: Rudall, P. J., Cribb, P. J., Cutler, D. F., and Humphries, C. J., (Eds.), Monocotyledons: Systematics and Evolution. Royal Botanic Gardens, Kew, Vol. 1, pp. 157–168.

Rudall, P. J., Furness, C. A., Chase, M. W., and Fay, M. F., 1997. Microsporogenesis and pollen sulcus type in Asparagales (Lilianae). Can. J. Bot. 75, 408–430.

Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M., 1996. Phylogenetic inference. In: Hillis, D. M., Moritz, C., Mable, B. K., (Eds.), Molecular Systematics, Second edition. Sinauer Associates, Sunderland, MA, pp. 407–543.

Takhtajan, A., 1980. Outline of the classification of flowering plants (Magnoliophyta). Bot. Rev. 46, 225–359.

Takhtajan, A., 1997. Diversity and classification of flowering plants. Columbia University Press, New York, pp. 472–526.

Thorne, R. F., 1983. Proposed new realignments in the angiosperms. Nordic. J. Botany. 3, 85–117.

Thorne, R. F., 1992. An updated classification of the flowering plants. Aliso 13, 365–389.

~~ CHAPTER 2 ~~

PARALLEL LOSS OF AN INTRON IN SEVERAL CLOSELY RELATED MONOCOTS (ASPARAGALES: ASPHODELACEAE AND HEMEROCALLIDACEAE) FROM A SLOWLY EVOLVING PLASTID GENE

Introduction

The plastid (chloroplast) genome shows extensive conservation in size, structure, composition and rate of nucleotide substitution across a broad spectrum of land plants (Palmer 1991; Clegg *et al.* 1994). However, its composition is not static; there is a general trend towards the loss of genetic information over time by gene deletion (through transfer to the nuclear genome or complete loss of function) and intron loss (Palmer 1991; Martin *et al.* 1998). Large structural mutations, such as intron losses, are often considered to be reliable phylogenetic markers, because of their relatively infrequent evolutionary occurrence (e.g., Downie *et al.* 1991; Downie and Palmer 1992). The stability of plastid introns in the land plants, and their lack of lateral spread among or within genomes and organisms (in contrast to some mitochondrial and bacteria; e.g., Cho *et al.* 1998; Dai and

Zimmerly 2002) may result from a lack of enzymatic machinery for intron mobility in this organelle (Bonen and Vogel 2001).

Intron loss is thought to result from reverse-transcription of a processed RNA transcript, with an intronless cDNA replacing the intron-bearing version of a gene by homologous recombination (see, for example, Palmer *et al.* 1991; Bonen and Vogel 2001). There is nothing about this molecular process to prevent it from being repeated multiple times in different lineages for any particular intron. However, convergent losses of introns are a relatively infrequent feature of plastid genome evolution at a broad phylogenetic scale in the flowering plants. Most homoplastic intron losses take place in relatively distantly related groups. They have only been reported in the angiosperms for introns of the plastid genes *rpl2* (Downie *et al.* 1991), *rpoC1* (Downie *et al.* 1996), *rpl16* (Downie and Palmer 1992; Campagna and Downie 1998), and the *cis*-spliced intron in the *trans*-spliced gene *rps12* (Hoot and Palmer 1994; Freyer *et al.* 1995). This general rarity means that intron loss may be a more reliable phylogenetic marker than gene loss (Doyle *et al.* 1995). We consider these literature reports to be instances of evolutionary convergence rather than parallelism, although these two related terms stand at ends of a spectrum of homoplasy, and the distinction between them will not always be completely clear. Nonetheless, convergent losses are considered to be relatively clear markers of relationship within the different groups in which they are found (e.g., Downie *et al.* 1991, 1996; Doyle *et al.* 1995).

During a DNA sequence-based investigation of higher-order relationships in the large monocot order Asparagales (Chapter 3), a large group of 29 families that includes the orchids, irises, and onions (APG 1998), we encountered instances of intronless versions of the *3'-rps12* locus. The locus is normally situated in the plastid inverted repeat (IR), an extremely slowly evolving portion of the plastid genome (Wolfe *et al.* 1987; Goremykin *et al.* 1996; Graham and Olmstead 2000a). The *rps12* gene product (plastid ribosomal protein S12) is produced from two transcripts with disjunct locations in the plastid genome. The 5'- and 3'- end transcripts are *trans*-spliced together, and *cis*-splicing removes the intron from the 3'- end transcription fragment (Koller *et al.* 1987; Zaita *et al.* 1987; Hildebrand *et al.* 1988).

We surveyed for the occurrence of the *cis*-spliced intron in the *3'-rps12* locus across a large sampling of taxa, and then examined evolutionary changes in its status (present vs. absent) in the context of a DNA sequence-based phylogenetic study of the order Asparagales (Fay *et al.* 2000, updated here). We reconstructed the evolution of intron status using a range of optimization schemes on an optimal tree and several suboptimal trees that were not significantly worse explanations of the DNA sequence data.

These reconstructions suggest that *3'-rps12* intron loss likely occurred more than once among two closely related lineages in the order: Asphodelaceae,the aloe family, and Hemerocallidaceae, the day lily family. .

Materials and Methods

Status of the 3'-rps12 *intron.* Genomic DNAs were extracted using the basic methodology of Doyle and Doyle (1987, see Fay *et al.* 2000 for further details). We determined the presence or absence of the plastid *3'-rps12* intron by examining the size of DNA amplification products generated using intron-flanking primer sites in the operon for *3'-rps12* and *rps*7 (primers 1F and 6R; Fig. 2.1). Taxa assessed for *3'-rps12* intron status in Asparagales and outside the order are listed in Table 2.1 and the Appendix, respectively (family circumscriptions in the flowering plants follow APG 1998, unless otherwise noted). The expected lengths of amplification products with and without the intron are ca. 1.3 kb and 0.75 kb respectively. For comparison, the corresponding sizes from *Nicotiana tabacum* L. (GenBank accession number Z00044) would be 1,284 bp and 748 bp. Part of the region amplified is the 3'-*rps*12–*rps7* intergenic spacer (Fig. 2.1). However, the estimated length variation in this additional noncoding region is small (standard deviation in length < 20 bp across a range of seed plants; Graham and Olmstead 2000a). For a subset of taxa, intron status was confirmed by sequencing the entire intron region (Table 2.1, Appendix). Amplification and sequencing of this region were carried out with primers and protocols outlined in Graham & Olmstead (2000a). For these taxa, intron/exon boundary sequences follow standard plastid group II intron junction motifs, and the intron is located in precisely the same position (data not shown). Intron deletion was precise and complete in all cases examined, in line with other studies (Downie *et al.* 1991, 1996, 1998; Hoot *et al.* 1994; Lai *et al.* 1997).

Phylogenetic analysis of Asparagales based on multiple plastid regions. A subset of taxa examined in Asparagales lack the *3'-rps12* intron. To place this putative loss in a phylogenetic context, we re-analyzed the data set of Fay *et al.* (2000), who examined DNA sequences from several regions in the plastid genome (*atpB*, *rbcL*, *trnL-F*) for 64 species representing 28 families of Asparagales and 14 outgroup taxa. We added four Hemerocallidaceae (*Corynotheca micrantha* (Lind.) Druce, *Pasithea coerulea* (Ruiz & Pavon) D. Don., *Stypandra glauca* R. Br., and *Tricoryne* R. Br. sp.) to their DNA-sequence matrix to improve taxon representation in this family, for which a subset of taxa lack the intron (see Results). Source information is provided in Table 2.1 and Fay *et al.*

(2000). The primers used for DNA amplification and sequencing *rbcL*, *trnL-F* and *atpB* are those of Zurawski *et al.* (1984) and Muasya *et al.* (1998), Taberlet *et al.* (1991), and Hoot *et al.* (1995), respectively. Protocols follow Graham and Olmstead (2000a). The following regions are included for the four additional taxa (GenBank accession numbers are noted after each region): *Corynotheca micrantha* – *atpB* (AY125895), *rbcL* (AY125899); *Pasithea coerulea* – *atpB* (AY125896), *rbcL* (Z77305), *trnL-F* (AY125900); *Stypandra glauca* – *atpB* (AY125897), *rbcL* (Z77307), *trnL-F* (AY125901); *Tricoryne* sp. – *atpB* (AY125898), *rbcL* (Z77308), *trnL-F* (AY125902). The *rbcL* data for *Pasithea* D. Don., *Stypandra* R. Br. and *Tricoryne* R. Br. were published previously (Rudall *et al.* 1997); the other sequences are new here.

The included portion of the *trnL-F* region consists of part of the plastid *trnL*(UAA) gene (partial exon 1, intron, and exon 2), the intergenic spacer region between *trnL*(UAA) and *trnF*(GAA), and a small portion of *trnF*(GAA). Short sections of the noncoding portions for *trnL-F* were difficult to align, and were omitted from phylogenetic analysis following exclusion sets used by Fay *et al.* (2000). We were unable to obtain sequence data for *trnL-F* for *Corynotheca* F. Muell. *ex* Benth., and obtained only partial data for *Tricoryne* (mostly limited to the intergenic spacer region). The unsequenced portions for these and other taxa were coded as "missing data" for affected taxa; several short portions at the ends of *atpB* and *rbcL* were also excluded (see Fay *et al.* 2000 for details). The addition of the new sequences to the matrix did not require the introduction of any new alignment gaps. There were 3,551 aligned nucleotides per taxon, corresponding to 3,331 unaligned nucleotides (reference taxon = *Xanthorrhoea resinosa* Pers.).

We performed maximum parsimony (MP) searches on the matrix using PAUP* version 4.0b10 (Swofford 2002). All characters and character-state changes were equally weighted, and heuristic searches were conducted using tree-bisection-reconnection (TBR) branch swapping, with 100 random addition replicates. A bootstrap analysis (Felsenstein 1985) was performed using the same search conditions, except that one random addition replicate was used for each of 100 bootstrap replicates. No tree limit was set for each bootstrap replicate.

Maximum likelihood analysis. The maximum parsimony heuristic search indicated the existence of two distinct, but closely related lineages within Asparagales that lack the *3'-rps12* intron (see Results). To facilitate further exploration of this finding using likelihood-based analysis, we reduced the plastid DNA sequence data set to nine exemplar taxa (*Aloe* L., *Asphodelus* L., *Dianella* Lam., *Hemerocallis* L., *Kniphofia* Moench, *Pasithea*, *Stypandra*, *Tricoryne* and *Xanthorrhoea*) that span the diversity of lineages in a large clade of three families (Asphodelaceae, Hemerocallidaceae and Xanthorrhoeaceae; see below) where some taxa lack the *3'-rps12* intron (see Results), together with three outgroup taxa (*Asparagus* from Asparagaceae, *Xeronema* Brongn. and Gris from Xeronemataceae, and *Gladiolus* L. from Iridaceae). All of the chosen taxa have complete or nearly complete sequences available for the plastid regions examined here: two taxa from Hemerocallidaceae that lack the *3'-rps12* intron (*Caesia* R. Br. and *Corynotheca*) were excluded because they have not been sequenced for the *trnL-F* region.

We performed two ML-based searches with this taxon subset. One was a straightforward heuristic tree search. We refer to this as the "unconstrained" search, and the tree that results from it (see Results) as the "unconstrained tree." A second search was performed with taxa constrained into two partitions that reflect each taxon's *3'-rps12* intron status. This constrained search enforces an assumption that intron absence among closely related taxa is a strong indicator of monophyly. We refer to the suboptimal tree that resulted from this search as the "constrained tree." Intron status was generally determined with one species per genus. Assessed taxa were not always conspecific with those in the DNA-sequence analysis (cf. Table 1 in Fay *et al.* 2000, and Table 2.1 and the Appendix here). No species of *Gladiolus* were examined for intron status, and so the constrained search was performed using the "Backbone constraints" function in PAUP* to permit this taxon to float freely during tree searching. *Asphodelus* had different species examined for the intron survey and the DNA sequence analysis (*Asphodelus albus* Willd. and *A. aestivus* Brot., respectively); three "composite" taxa in the DNA sequence matrix (*Dianella, Hemerocallis* and *Xanthorrhoea*; see Fay *et al.* 2000) had one component species that was also examined here for intron status (Table 2.1); the remaining samples examined for intron loss for the ML analysis were conspecific with those in the DNA-sequence matrix.

A substitution model was chosen prior to ML-based tree searching using the likelihood ratio test (LRT), as described in Huelsenbeck and Crandall (1997). We added an additional substitution model (general time-reversible, GTR; Lanave *et al.* 1984; Tavaré 1986; Barry and Hartigan 1987; Rodríguez *et al.* 1990) to their hierarchy of tests, and also did not test for a molecular clock. The model tree used for the LRT was the single best MP tree found for these twelve taxa; the likelihood model chosen was GTR + Γ (Table 2.2). Substitution model parameters [the substitution rate matrix (R-matrix) parameters, the base frequencies, and the Γ–shape parameter, α] were estimated during tree inference for the constrained and unconstrained tree searches. The likelihood tree searches were conducted using TBR branch swapping.

SOWH test of alternative trees. Character reconstructions were performed on five trees in total: two trees generated from tree searches (the unconstrained and constrained trees), and three additional suboptimal trees that we derived by making minor alterations in the topology of the unconstrained tree. These three "manipulated trees" were considered because they are potentially consistent with radically different character reconstructions than those inferred on the unconstrained tree, but do not appear to be substantially less optimal than this tree (see below).

We determined the significance of the increase in length of each of the four suboptimal ML trees (the constrained tree and the three manipulated trees) relative to the optimal (unconstrained) tree using a parametric bootstrapping method described by Swofford *et al.* (1996). The SOWH test (the name is an acronym of the authors' surnames) focuses on a pair of trees. The null hypothesis is that the less optimal of the pair is actually the correct tree, but that the more optimal tree was obtained in practice due to the effects of sampling error on finite data.

We implemented their test by generating 1,000 data sets as large as the real data set by Monte Carlo simulation on a given null topology (the suboptimal one according to the real data), using Seq-Gen v1.1 (Rambaut and Grassly 1997). The branch lengths and substitution model parameters for the Monte Carlo simulations were estimated from the null tree using the real plastid data, under the appropriate DNA substitution model (i.e., GTR + Γ).

The multiple simulated data sets for a particular null hypothesis were then assessed in batch mode in PAUP*. For each simulated data set, the ML tree score was re-calculated using the null topology, and an unconstrained heuristic search was performed, using re-calculated GTR + Γ substitution model parameters determined for that simulated data set on the null tree. Our SOWH tests thus use "partial" likelihood optimization and correspond to the "posPpud" variant described in Goldman *et al.* (2000). The same basic heuristic search conditions were used as with the real data to find the shortest tree or trees (the "alternative hypothesis") for each simulated data set. Tree searches used TBR branch swapping. We repeated this process for all 1,000 simulated data sets, recording tree scores for null and alternative hypotheses in each case. The distribution of the differences in likelihood scores between the null and alternative hypotheses (trees) for each simulated data set can be used to determine the significance of the corresponding difference, δ, for the unconstrained vs. suboptimal tree scores from the real data. The SOWH test was repeated for the four suboptimal topologies under consideration, by comparing each, in turn, to the optimal, unconstrained tree inferred from the real plastid data.

Reconstruction of the evolution of intron status using plastid-based phylogenies. Ancestral state reconstructions were performed using MacClade 4.0 (Maddison and Maddison 2001). Two main types of character transformation scheme were considered for these maximum parsimony reconstructions. One scheme, which we refer to as "flat weighting," confers equal character-state transformation weights on intron gain versus loss. For binary characters, flat-weights can be implemented using either unordered parsimony (Fitch 1971; Hartigan 1973) also known as Fitch parsimony, or ordered parsimony (Farris 1970; Swofford and Maddison 1987), also known as Wagner parsimony (see Swofford *et al.* 1996; pp. 422-424 for a description of the distinction between these two transformation schemes). We used the former algorithm in MacClade, but the two algorithms give equivalent results, as character ordering is not applicable to binary data. A flat weighting scheme can be a strong assumption to make (e.g., Fitch 1984; Swofford *et al.* 1996; p. 423), particularly when previously published evidence suggests that there may be a strong bias in the gain or loss of a particular character state. In this situation some authors (e.g., Maddison and Maddison 1992, pp. 70-71; Kohn *et al.* 1996; Omland 1997, 1999) advise that biased weighting schemes should be considered that better reflect available biological knowledge. We addressed this concern using simple step matrices that confer mild biases against the evolutionary gain of the *3'-rps12* intron (referred to as "relaxed Dollo" schemes by Swofford *et al.* 1996; p. 421).

Results

Intron survey. Of the monocots examined here, the *3'-rps12* intron is missing only in Asphodelaceae and some members of Hemerocallidaceae (Table 2.1; Appendix 2.1; Figs. 2.2, 2.3). We surveyed one or more representatives for all genera in the former family (of 15 genera of Asphodelaceae accepted by Smith and Van Wyk 1998), except *Chortolirion* A. Berger and *Lomatophyllum* Willd. The intron is absent in all Asphodelaceae examined. We also scored the intron status of one or more representatives of 13 of the ca. 19-20 genera currently recognized in Hemerocallidaceae (see APG 1998; Clifford and Conran 1998; Clifford *et al.* 1998). The intron is present in all taxa examined in Hemerocallidaceae, except for *Hemerocallis*, *Simethis* Kunth. and all sampled taxa in the tribe Johnsonieae Benth. The latter tribe is also recognized at the family level as Johnsoniaceae J. P. Lotsy (the current circumscription of which follows Clifford and Conran 1998; see also Chase *et al.* 1996). Johnsoniaceae are accepted as a synonym of Hemerocallidaceae by APG (1998).

Asphodelaceae and Hemerocallidaceae are part of a well-supported clade that includes Xanthorrhoeaceae (Chase *et al.* 2000a,b; Fay *et al.* 2000; Fig. 2.3), a monogeneric family with ca. 30 species. We sampled three species of *Xanthorrhoea* here (Table 2.1); all possess the intron. Hemerocallidaceae as circumscribed by APG (1998) excludes the monogeneric family Xeronemataceae (see Chase *et al.* 1996). The single sampled species of *Xeronema* possesses the intron, as do all representatives of the other families surveyed in Asparagales (Table 2.1; Fig. 2.3). Examples of DNA amplifications of taxa in Asparagales with and without the intron are shown in Fig. 2.2. Table 2.1 and the Appendix provide a complete list of taxa surveyed.

Phylogenetic context for intron loss. We inferred essentially the same set of phylogenetic relationships of Asparagales as Fay *et al.* (2000), using our modification of their DNA sequence data matrix. Differences in tree topology involve the local positions of Aphyllanthaceae, Doryanthaceae, Hyacinthaceae, Ixioliriaceae, Tecophilaeaceae and Themidaceae. *Xeronema* is the sister group of a large clade of taxa that includes Asphodelaceae-Hemerocallidaceae-Xanthorrhoeaceae and the other “higher Asparagales” of Chase *et al.* (1995). Within the former clade, the family Xanthorrhoeaceae has slightly different positions in the four MP trees (as it did in Fay *et al.* 2000). In two of the four trees, Xanthorrhoeaceae are the sister group of Asphodelaceae-Hemerocallidaceae (Fig. 2.3); in the other two the family is the sister group of Asphodelaceae only; neither position is well supported by bootstrap analysis (< 50% support in both cases). Bootstrap support for much of the backbone of relationships in Asphodelaceae-Hemerocallidaceae-Xanthorrhoeaceae is also weak, although the clade formed by the three families is robustly supported, and Asphodelaceae and Hemerocallidaceae each have robust bootstrap support (100% and 91%, respectively).

Relationships among the constituent genera of Asphodelaceae and Hemerocallidaceae are otherwise identical to those shown in Fig. 1 of Fay *et al*. (2000), when the four additional genera sampled here are pruned from our tree (i.e., *Corynotheca*, *Pasithea*, *Stypandra*, and *Tricoryne*). Within Hemerocallidaceae, *Pasithea* is weakly supported as the sister group of the taxa included here. *Stypandra* is strongly supported as the sister group of *Dianella*, an arrangement noted previously by Rudall *et al*. (1997), although with only poor support there. *Corynotheca* is strongly supported as the sister group of *Caesia*. These two taxa and *Tricoryne* are in turn resolved as a poorly supported terminal clade in Hemerocallidaceae. This clade corresponds to Johnsoniaceae, and so the MP trees support the synonymy of this family with Hemerocallidaceae (Table 2.1; six genera of Johnsoniaceae were sampled for intron status; the remaining two genera, *Arnocrinum* Lehm. and *Hodgsoniola* F. Muell., were not surveyed here due to lack of material). However, the weakness observed along most of the backbone of relationships within Hemerocallidaceae means that a simple sister-group relationship between Hemerocallidaceae and Johnsoniaceae cannot be ruled out.

Reconstruction of the evolution of 3'-rps12 *status in Asparagales*. The optimal ML tree inferred from the reduced taxon set (Fig. 2.4) has the same topology as the MP tree shown in Fig. 2.3, when the latter tree is pruned to this taxon set. The most straightforward interpretation of this tree with regards to the *3'-rps12* intron is that there were two parallel losses of the *3'-rps12* intron in Asparagales (Fig. 2.4a). When flat weights are used to map intron status onto the unconstrained ML tree, independent losses are inferred in two closely related lineages; Asphodelaceae and *Hemerocallis*/ *Tricoryne* of Hemerocallidaceae (Fig. 2.4a). By this interpretation, one intron loss may act as a synapomorphy of at least part of Hemerocallidaceae. A similar reconstruction results from mapping intron loss on the main MP trees (not shown); two independent intron loss events are inferred to have occurred, one in the lineage leading to the sampled taxa in Asphodelaceae, and one in the line leading to at least five genera of Hemerocallidaceae (*Caesia*, *Corynotheca*, *Hemerocallis*, *Simethis* and *Tricoryne*).

An alternative interpretation is possible: that the intron loss should be recognized as a synapomorphy for the monophyly of the group of taxa in Asphodelaceae and Hemerocallidaceae that lack it, and therefore that phylogenetic inferences for these taxa based on DNA sequence data are incorrect. That does not appear to be the case. The relatively strong support for the monophyly of Hemerocallidaceae (91%) conflicts with this hypothesis, and the best tree constraining the monophyly of the intronless taxa from Asphodelaceae and Hemerocallidaceae is significantly longer than the best unconstrained likelihood tree, according to the SOWH test ($P < 0.05$; Fig. 2.4b). This constrained hypothesis, had it been plausible, would also clearly have implied a single loss of the intron in Asparagales (Fig. 2.4b).

There are subtler ways in which a reconstruction using flat weights could have indicated a single loss of the intron in Asparagales. Plausible phylogenetic scenarios that place *Hemerocallis* and *Tricoryne* at or near the base of Hemerocallidaceae can also result in a single loss (Fig. 2.4c). With *Hemerocallis* forced to the base of Hemerocallidaceae, and *Tricoryne* to the next basal split, or vice versa, a single loss and a subsequent secondary gain are inferred (Fig. 2.4c; left-hand and middle trees). When both taxa are forced to form a basal clade in Hemerocallidaceae, this yields an equivocal reconstruction, in which there are either two losses, or again, a single loss and a later, secondary gain of the intron (Fig. 2.4c; right-hand tree). There is poor bootstrap support for the backbone of relationships within Hemerocallidaceae, and so none of these tree topologies can be ruled out in the current study. The SOWH tests confirm this: none of the three manipulated trees are significantly longer than the best tree (Fig. 2.4). We must therefore entertain all three trees shown in Fig. 2.4c as reasonable topological scenarios.

Although a single loss and subsequent secondary gain of the *3'-rps12* intron cannot be ruled out under reconstructions that apply flat weights, it is not difficult to tip the balance to find weighting schemes that result in multiple parallel losses with no regains of the intron. We performed a simple sensitivity analysis to examine a range of "relaxed Dollo" character-state transformation schemes that weight slightly against the gain of introns. All schemes that weight against secondary intron gain by a marginally greater than two-fold cost differential result in three parallel losses (and no secondary intron gain) for the first two trees shown in Fig. 2.4c. One of these cases is illustrated in Fig. 2.5. For the third manipulated tree (Fig. 2.4c), the bias against secondary intron gain needs to be only marginally more costly than even weights to yield a reconstruction with two losses and no secondary gain of the intron (not shown); DELTRAN resolution of the equivocal branch yields this same evolutionary reconstruction.

Discussion

Reconstruction of 3'-rps12 *intron loss*. The various reconstructions support two major classes of evolutionary change in the *3'-rps12* intron in Asparagales (Figs. 2.4, 2.5): there were either multiple losses, or a single loss with a secondary gain of the intron. Some of the suboptimal trees considered are not significantly longer than the optimal tree according to the SOWH test, and yield the latter class of reconstruction ("loss, then gain") when flat weights are used to reconstruct evolutionary change for this intron (Fig. 2.4c).

If true, scenarios with secondary gain of the *3'-rps12* intron would be very surprising, for several reasons. First, no secondary gain of a plastid group II intron has ever been reported in the land plants (Palmer 1991; Downie *et al.* 1998). Indeed, the enzymatic machinery for plastid intron mobility may be lacking in land-plant plastids (Bonen and Vogel 2001). Second, the taxa where a secondary gain would be necessary (*Dianella*, *Pasithea* and *Stypandra*; Fig. 2.4c) all have introns located in precisely the same position observed in other land plants (for reference, exons flanking the intron end and begin at bp 232 and 769 respectively in tobacco; GenBank accession number Z00044). A precise secondary re-gain would most likely require "retrohoming" ; a phenomenon documented in bacteria and fungi, but not previously reported for plastid genomes (e.g., Dai and Zimmerly 2002). These considerations lead us to reject some evolutionary scenarios implied using flat-weighted parsimony reconstructions on plausible suboptimal trees (Fig. 2.4c), to favor instead reconstructions based on mildly skewed weights. These reconstructions imply only parallel losses (perhaps three in total; Fig. 2.5). A sensitivity analysis demonstrates that the skew in weights required to unequivocally reconstruct a single loss on all plausible trees considered here is mild: slightly greater than two-fold, or barely skewed at all if *Hemerocallis*/ *Tricoryne* constitute a clade at the base of Hemerocallidaceae (Results; Fig. 2.4c).

The idea that *3'-rps12* intron loss in Asparagales might be a strong marker for the monophyly of all of the taxa that lack it is rejected by the SOWH test (Fig. 2.4b). Intron loss is not a synapomorphy for these taxa (taken collectively), and it therefore does not contradict the monophyly of Hemerocallidaceae. Intron losses may still support the monophyly of subsets of taxa in Asparagales, and indeed one loss appears to act as a synapomorphy for the inclusion of Johnsoniaceae within Hemerocallidaceae (as tribe Johnsonieae). *Hemerocallis*, *Simethis* and all Johnsonieae sampled here lack the intron (Table 2.1; Fig. 2.3). However, since our results show that parallel loss is a reality among closely related taxa in Asparagales, we should not completely discount the idea that there might also have been multiple homoplastic losses of the intron within Hemerocallidaceae (e.g., Fig. 2.5); further clarification of the phylogenetic status of Johnsoniaceae within Hemerocallidaceae must await new phylogenetic data (both genes and taxa).

3'-rps12 *intron loss in the embryophytes.* The *3'-rps12* intron is present in almost all land plants examined to date (Appendix 1; Graham and Olmstead 2000a, b; Graham *et al.* 2000; S. W. Graham *et al.* unpubl. data). Previously it has been reported as missing in the plastid genomes of three groups of eudicots: the parasitic plant *Cuscuta europaea* L. (Convolvulaceae; Freyer *et al.* 1995), three of 37 species examined in *Anemone* L. (Ranunculaceae; Hoot and Palmer 1994), and two of four species examined in Fabaceae (Appendix 1). The species of *Anemone* lacking the intron form a small and well-supported clade within Ranunculaceae (Hoot and Palmer 1994). In Fabaceae, the intron is absent in *Pisum sativum* L. of tribe Vicieae (see Graham *et al.* 2000) and *Medicago truncatula* Gaertn. of tribe Trifolieae, but is present in *Lotus japonicus* (Regel) K. Larsen of tribe Loteae and *Glycine max* (L.) Merr. of tribe Phaseoleae. Three of these four legumes are closely related (*Lotus* L., *Medicago* L. and *Pisum* L.) and the two more closely related taxa lack the intron (see Kajita *et al.* 2001), which suggests that the loss of this intron may have affected a relatively limited subset of Fabaceae (although see below). As part of a phylogenetic study of the pteridophytes, a case of *3'-rps12* intron loss has also been encountered in *Equisetum laevigatum* A. Br. (H. S. Rai *et al*. unpubl. data). Equisetaceae, Convolvulaceae, Fabaceae, and Ranunculaceae are all distantly related to each other (Ranunculaceae is a basal eudicot; Convolvulaceae and Fabaceae belong to euasterid I and eurosid I, respectively, in the eudicots; see e.g., APG 1998).

The intron is otherwise uniformly present in other surveyed vascular plants (Appendix 1; Graham *et al*. 2000; S. W. Graham *et al*. unpubl. data). There were therefore a minimum of four additional convergent losses of the *3'-rps12* intron in the vascular plants, apart from the losses found here. For both cases in the core eudicots, there may be exceptional circumstances associated with the intron loss. *Cuscuta europaea* is a holoparasitic flowering plant; holoparasitism (and associated achlorophylly) are often accompanied by substantial plastid genome disruption (see Nickrent *et al*. 1999). Although Fabaceae are non-parasitic, structural rearrangements in the plastid genome also seem to be the rule rather than the exception in this family. A subset of legumes have undergone substantial rearrangements in gene order (through inversions and reduction in the extent of the plastid IR region; Palmer *et al*. 1987, 1988), and there are three reported cases in the family of parallel intron loss. The *rpo*C1 intron has probably been lost multiple times within *Medicago* (Downie *et al*. 1998) and the *rpl*2 intron has been lost repeatedly, in tribe Phaseloeae and relatives in subfamily Faboideae (Doyle *et al.* 1995) and perhaps also in *Bauhinia* L., subfamily Caesalpinioideae (Lai *et al.* 1997).

Genome-level phenomena and implications for intron loss. A major conclusion of this study is that parallel losses of the *3'-rps12* intron are likely to have occurred in two closely related lineages of Asparagales (Figs. 2.4a, 2.5). However, this parallel loss is fairly surprising. These monocot taxa are not parasitic (the most closely related achlorophyllous plants are in Orchidaceae; Dressler 1993) and whereas Asphodelaceae and Hemerocallidaceae have not been studied previously for plastid genome organization, a range of other plastid genes and noncoding regions in *Hemerocallis littorea* Makino and *Asphodelus albus* do not appear to have a particularly elevated rate of molecular evolution (for *atpB*, *ndhF*, *ndhB*, *rbcL*, *rpl2*, *rps7*, ten photosystem II genes and associated noncoding regions; Chapter 3).

The *3'-rps12* locus is generally located in the plastid IR region. This region evolves extremely slowly at the nucleotide level (Wolfe *et al.* 1987; Goremykin *et al.* 1996; Graham and Olmstead 2000a). It is not clear if the rate of loss of the *3'-rps12* intron is in any way retarded in relation to introns from the more rapidly evolving regions of the plastid genome. The relative rates of structural vs. nucleotide substitution processes may have nothing to do with each other; the biochemical processes involved are likely to be unrelated. Our argument that elevated rates of structural mutations in some legumes may be associated with elevated rates of intron loss may therefore be faulty for the same reason. Nonetheless, it is clear that the rate of microstructural changes (indels) is also substantially lower in the IR region (Graham *et al.* 2000). If the rates of indel and nucleotide mutation are evolutionarily correlated in the IR, then the other mutational processes may be too. At present however, there are not enough cases of intron loss and other major structural mutations to do more than raise this as a possibility.

Lai *et al.* (1997) noted that multiple *rpl*2 intron losses in *Bauhinia* may represent parallel fixations of an intron presence/ absence polymorphism that was retained across several closely spaced speciation events (at the biochemical level there would only have been a single intron loss). However, if differential sorting of ancestral intron loss polymorphisms did occur in Asphodelaceae and Hemerocallidaceae, the intronless taxa should also be (misleadingly) linked as closest relatives on the DNA-sequence based tree, because the tree on which we inferred intron loss is based on exactly the same frozen linkage group (the plastid genome) from which the intron is lost. As they are not inferred to be part of the same clade (Figs. 2.3-2.5), this sorting mechanism cannot explain the observed phylogenetic distribution situation of intronless taxa in Asparagales.

Ancestral state reconstruction. Our exploration of character reconstruction on optimal and suboptimal trees was triggered by our initial surprise about the observed distribution of intron absence in the context of what we understood about Asparagales phylogeny and taxonomy, the overall rarity of group II intron loss in the land plants, and the apparent improbability of regaining them, once lost. We examined the loss in a parsimony framework. Several likelihood-based approaches (e.g., Sanderson 1993; Pagel 1994, 1997; Mooers and Schluter 1999) have been also used for reconstructing ancestral states, and for assessing whether one-rate and two-rate models of character evolution are significantly different (flat-weights in parsimony correspond to the one rate model; uneven parsimony state-transformation costs correspond to two-rate models). A Bayesian framework would allow us to better quantify the degree of surprise of an observation in relation to our prior knowledge and the likelihood of observed character distributions (e.g., Huelsenbeck *et al.* 2002), in the context of a model of evolution (here the "model" would include the tree and a one-rate or two-rate state transformation model). Likelihood-based approaches generally assume that the same transformation model applies across the entire tree (e.g., Lewis 2001). There is good reason to believe that this might not hold for intron loss reconstructions, as some groups of organisms have elevated rates of loss (see above). The sensitivity of parsimony-based reconstructions to this phenomenon is also unknown.

Our parsimony-based reconstructions provide lucid summaries of the possible scenarios for the evolution of intron status. Consideration of some surprising reconstructions on plausible phylogenies of Asparagales spurred us to apply a sensitivity analysis (Donoghue and Ackerly, 1996) to determine how heavily we would have to weight against particular scenarios (such as "intron re-gain") to yield alternative reconstructions. Evidently the weights required are minimal. The use of skewed parsimony weighting schemes to reconstruct the evolutionary history of intron loss is not unique to this study and is implicit or explicit in other studies that found multiple intron loss in the plastid genome (Downie *et al.* 1991, 1996, 1998; Hoot *et al.* 1994; Doyle *et al.* 1995; Lai *et al.* 1997), and multiple loss of group II introns in flowering-plant mitochondrial genomes (Joly *et al.* 2001).

It may seem circular to apply weights in this manner, but it has been argued repeatedly (and convincingly, we think) that it would be unwise to ignore external evidence that suggests that particular state-transformation schemes are more or less likely, when considering which type of parsimony scheme to apply to character reconstructions (e.g., Maddison and Maddison 1992, Kohn *et al.* 1996; Omland 1997). Swofford *et al.*

(1996, p. 423) also remind us that the use of equal weights is a strong *a priori* assumption. Finally, scenarios involving secondary gain of introns, although widely recognized as being highly improbable, must at least enter the arena of possibility when we think about the molecular biology and patterns of evolutionary change in group II introns in the plastid genome (e.g., Fig. 2.4c).

Caveat. We were not always able to score exactly the same taxa in our intron survey as in the DNA-sequence based study of Asparagales phylogeny; we therefore assumed that the intron status in a species is representative of its congener on the DNA-sequence based phylogeny. It is possible that this will result in misscoring of the status of some taxa used in the reconstructions of intron loss (Figs. 2.4, 2.5), due to unobserved parallel intron loss within a genus, or the lack of monophyly of some genera. (Note, however, that where did sample multiple taxa per genus they had the same intron status; Table 2.1.) If we did miss additional examples of intron loss, the effect should only be to increase the number of instances of intron loss inferred.

Conclusion

Multiple cases of repeated losses of plastid group II introns have been reported among distantly related taxa, but instances of parallel loss are much rarer, and were previously limited to taxa with elevated levels of genome rearrangement. The balance of evidence indicates that the *3'-rps12* intron was likely lost independently – at least twice – among very closely related lineages in the monocot order Asparagales. Because they are large structural changes, plastid intron losses caught the attention of systematists as potentially very reliable markers of evolutionary relationship. The homoplasy demonstrated here in *3'-rps12* intron loss is in line with evolutionary studies of other large structural mutations (e.g., Doyle *et al.* 1995; Graham and Olmstead 2000b) and reinforces the idea that no class of phylogenetic characters provides privileged phylogenetic information (e.g., Davis and Heywood 1973).

Table 2.1. Taxa examined for 3'-*rps*12 intron status in Asparagales.

Taxon	Source	Intron
Agapanthaceae		
Agapanthus coddii Leighton	Dept. Biol. Sci. Greenhouse (U. Alberta), ex Kirstenbosch Botanic Garden [1]	Present
Agapanthus africanus (L.) Hoffmgg.	M. A. McPherson 020420-1 (ALTA)	Present
Agavaceae		
Agave americana L.	Dept. Biol. Sci. Greenhouse (U. Alberta) [1]	Present
Yucca glauca Nutt.	[Voucherless field collection: Transect #6, Onefour, Alberta; coll. J. Addicott & D. Hurlburt (SWG 00121 DNA)]	Present [2]
Alliaceae		
Allium textile A. Nelson & J. F. MacBride	M. A. McPherson 990704-79 (ALTA)	Present [2]
Amaryllidaceae		
Amaryllis L.	Dept. Biol. Sci. Greenhouse (U. Alberta) [1]	Present
Narcissus elegans (Haw.) Spach	Barrett 1434 (TRT)	Present [2]
Anthericaceae		
Chlorophytum comosum (Thunb.) Jacques	M. A. McPherson 000321-1 (ALTA) [3]	Present [2]
Aphyllanthaceae		
Aphyllanthes monspeliensis L.	S. W. Graham 00-04-08 (ALTA)	Present [2]
Asparagaceae		
Asparagus officinalis L.	M. A. McPherson 010819-2 (ALTA)	Present [2]
Asphodelaceae		
Aloe purpurea Lam.	M. W. Chase 694 (K)	Absent
Aloe aculeata Pole Evans	Dept. Biol. Sci. Greenhouse (U. Alberta), ex Kirstenbosch Botanic Garden [1]	Absent
Aloe vera (L.) N. L. Burman	M. A. McPherson 010220-2 (ALTA)	Absent [2]
Asphodeline lutea (L.) Reichb.	UCI Arb 3440	Absent
Asphodelus albus Willd.	Lawrence Harder 1-000430 (ALTA)	Absent [2]
Astroloba foliolosa (Haw.) Uitewaal	Chase 684 (K)	Absent

Astroloba deltoidea (Hook. f.) Uitewaal	Ecology & Evol. Bio. Conservatory (U. Connecticut) [1]	Absent [2]
Bulbine frutescens (L.) Willd.	Ecology & Evol. Bio. Conservatory (U. Connecticut) [1]	Absent [2]
Bulbine succulenta Compton	UCI Arb 007174	Absent
Bulbinella cauda-felis Th. Dur. & Schinz	UCI Arb 359	Absent
Eremurus robustus (Regel) Regel	Dept. Biol. Sci. Greenhouse (U. Alberta) [1, 3]	Absent [2]
Eremurus sogdianus (Regel) Franch.	Popovet Vvedensky s.n. (ALTA)	Absent [2]
Eremurus spectabilis M. Bieb.	M. W. Chase 490 (K)	Absent
Gasteria verrucosa Duval	M. A. McPherson 010309-1 (ALTA) [3]	Absent [2]
Haworthia attenuata (Haw.) Haw.	Dept. Biol. Sci. Greenhouse (U. Alberta) [1, 3]	Absent [2]
Haworthia coarctata Haw.	M. W. Chase 3859 (K)	Absent
Haworthia turgida Haw.	M. A. McPherson 010309-3 (ALTA) [3]	Absent [2]
Jodrellia macrocarpa Baijnath	M. W. Chase 3941 (K)	Absent
Kniphofia linearifolia Baker	Dept. Biol. Sci. Greenhouse (U. Alberta), ex Kirstenbosch Botanic Garden [1]	Absent
Kniphofia uvaria (L.) Hook.	M. W. Chase 120 (NCU)	Absent
Kniphofia uvaria (L.) Hook.	M. A. McPherson 010508-1 (ALTA) [3]	Absent [2]
Poellnitzia rubriflora (L. Bolus) Uitewaal	M. W. Chase 669 (K)	Absent
Trachyandra Kunth.	M. W. Chase 1027 (K)	Absent
Asteliaceae		
Milligania stylosa (Hook. f.) Benth.	M. W. Chase 511 (K)	Present
Astelia alpina R. Br.	M. W. Chase 1103 (NCU)	Present [2]
Blandfordiaceae		
Blandfordia punicea (Labill.) Sweet	M. W. Chase 519 (NCU)	Present [2]
Boryaceae		
Alania endlicheri Kunth	D. H. Vitt 27706 (ALTA)	Present [2]

Convallariaceae		
Beaucarnea recurvata Lem.[4]	M. A. McPherson 011019-1 (ALTA)	Present
Maianthemum racemosa (L.) Link	M. A. McPherson 990704-97 (ALTA)	Present[2]
Doryanthaceae		
Doryanthes excelsa Correa	M. W. Chase 188 (NCU)	Present
Doryanthes palmeri Hill ex. Benth.	M. W. Chase 2837 (K)	Present
Hemerocallidaceae		
Caesia contorta Th. Dur. & Schinz.[5]	Goldblatt 9406 (MO)	Absent
Corynotheca micrantha (Lind.) Druce[5]	M. W. Chase 2210 (K)	Absent
Dianella montana Blume	M. W. Chase 2153 (K)	Present[2]
Dianella nigra Colenso	S. W. Graham 02-03-05 (ALTA)	Present[2]
Hemerocallis littorea Makino	M. W. Chase 3833 (K)	Absent[2]
Hemerocallis L. sp.	M. A. McPherson 010821-1 (ALTA)[3]	Absent[2]
Hensmania chapmanii G. J. Keighery[5]	M. W. Chase 2212 (K)	Absent
Herpolirion novae-zelandiae Hook. f.	R. & E. Melville 64862/62 (K)	Present
Johnsonia pubescens Lindl.[5]	M. W. Chase 2213 (K)	Absent[2]
Pasithea coerulea (Ruiz & Pavon) D. Don.	M. W. Chase 512 (K)	Present[2]
Phormium tenax J. R. Forst. & G. Forst.	M. W. Chase 177 (NCU)	Present[2]
Phormium tenax J. R. Forst. & G. Forst.	M. A. McPherson 000612-3 (ALTA)	Present[2]
Simethis mattiazzii (Vandelli) Sacc.	S. Ortiz 19-IV-1995-1	Absent[2]
Stawellia dimorphantha F. Muell.[5]	P. J. Rudall s. n. (ex Microprop) (K)	Absent
Stypandra glauca R. Br.	A. J. Whalen & R. G. Coveny 16660 (K)	Present[2]
Tricoryne R. Br. sp.[5]	M. W. Chase 2220 (K)	Absent
Hyacinthaceae		
Ornithogalum longebracteatum Jacq.	M. A. McPherson 020624-1 (ALTA)	Present
Muscari comosum (L.) Miller	L. Harder 000419-1 (ALTA)	Present[2]

Hypoxidaceae		
Curculigo capitulata (Lour.) Kuntze	M. W. Chase 205 (NCU)	Present[2]
Rhodohypoxis baurii (Baker) Nel.	Ecology & Evol. Bio. Conservatory (U. Connecticut)[1]	Present
Iridaceae		
Aristea eckloni Baker	Dept. Biol. Sci. Greenhouse (U. Alberta), ex Kirstenbosch Botanic Garden[1]	Present
Iris missouriensis Nutt.	M. A. McPherson 000707-5a-7 (ALTA)	Present[2]
Sisyrinchium montanum Greene	M. A. McPherson 990704-71 (ALTA)	Present[2]
Ixioliriaceae		
Ixiolirion tataricum (Pall.) Herb. & Traub	M. W. Chase 489 (K)	Present[2]
Lanariaceae		
Lanaria lanata (L.) Druce	M. W. Chase 458 (NCU)	Present[2]
Laxmanniaceae		
Lomandra longifolia Labill.	D. H. Vitt 27411 (ALTA)	Present[2]
Orchidaceae		
Coelogyne cristata Lindl.	M. A. McPherson 010921-1 (ALTA)[3]	Present[2]
Cypripedium passerinum Richardson	M. A. McPherson 010722-6 (ALTA)	Present[2]
Amerorchis rotundifolia (Banks) Hultén	M. A. McPherson 010610-1 (ALTA)	Present[2]
Tecophilaeaceae		
Cyanastrum cordifolium Oliver	Graham & Barrett 2 (TRT)	Present[2]
Themidaceae		
Milla biflora Cav.	Aaron Rodriguez 634 (WIS)	Present
Muilla maritima S. Watson	JCP 98-028 (WIS)	Present[2]
Brodiaea minor (Benth.) S. Watson	JCP 98-068 (WIS)	Present

Xanthorrhoeaceae		
Xanthorrhoea australis R. Br.	Ecology & Evol. Bio. Conservatory (U. Connecticut) [1]	Present [2]
Xanthorrhoea reflexa D. A. Herbert	R. J. Bayer #WA - 94072 (ALTA)	Present
Xanthorrhoea resinosa Pers.	M. W. Chase 192 (NCU)	Present [2]
Xeronemataceae		
Xeronema callistemon W. R. B. Oliv.	M. W. Chase 653 (K)	Present [2]

[1] Voucherless greenhouse specimen.

[2] Intron status confirmed by DNA sequencing.

[3] Cultivar.

[4] Syn. *Nolina recurvata* (Lem.) Hemsl. (Fay *et al.* 2000; Bogler, 1998).

[5] Johnsoniaceae J.P. Lotsy (sensu Clifford and Conran, 1998).

Table 2.2. Likelihood ratio test (LRT) of different substitution models for a combined data set of three plastid regions (*atp*B, *rbc*L and *trn*L-F), on a maximum parsimony tree inferred from a subset of taxa in Asparagales [1].

Substitution model [2]	-ln likelihood	Comparison[1]	$-2 \ln \Lambda$ [3]	d.f.	P [4]
JC69	9811.09	-----	-----	-----	-----
F81	9758.34	JC69 vs. F81	105.50	3	< 0.01
HKY85	9594.44	F81 vs. HKY85	327.80	1	< 0.01
GTR	9562.58	HKY85 vs. GTR	63.72	4	< 0.01
GTR + Γ	9369.87	GTR vs. (GTR + Γ)	385.42	1	< 0.01
GTR + Γ + I	9367.07	(GTR + Γ) vs. (GTR + Γ + I)	5.6	1	NS

[1] Twelve taxa from Figure 2.3; see Figs. 2.4, 2.5.

[2] Abbreviations: JC69 = Jukes and Cantor 1969; F81 = Felsenstein 1981; HKY85 = Hasegawa et al. 1985; GTR = General time-reversible (Lanave et al. 1984; Tavaré 1986; Barry and Hartigan 1987; Rodríguez et al. 1990). Γ = Gamma; I = Proportion of invariable sites.

[3] Likelihood ratio test statistic.

[4] The α-level was adjusted using the Bonferroni correction for five tests; NS = not significant.

FIG. 2.1. Overview of the plastid 3'-*rps*12 and *rps*7 operon. The two primers indicated were used to survey for 3'-*rps*12 intron presence or absence. Graham and Olmstead (2000a) list primer sequences. Map lengths are based on *Nicotiana tabacum* (GenBank accession Z00044); a scale bar is indicated (primers not drawn to scale).

Figure 2.1

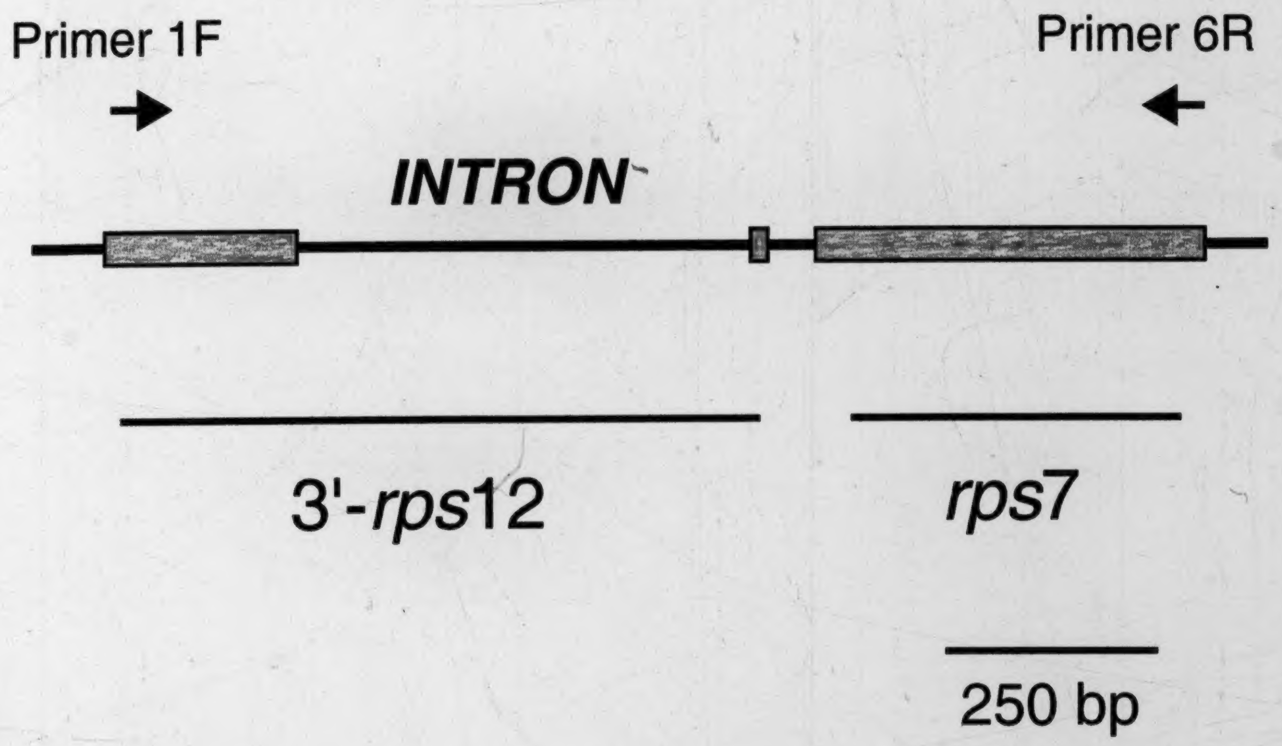

Primer 1F
Primer 6R
INTRON
3'-rps12
rps7
250 bp

FIG. 2.2. Ethidium-bromide stained agarose gel showing amplification products for a subset of taxa surveyed in Asparagales for 3'-*rps*12 intron presence/ absence. The primers used for amplification are shown in Fig. 2.1. Several marker band sizes are indicated (fragments smaller than 1500 bp decrease in size in 100 bp increments). Intron status: Intron present, lanes (1) *Amerorchis rotundifolia*, (2) *Xeronema callistemon*, (3) *Xanthorrhoea resinosa*, (4) *Dianella montana*, (5) *Phormium tenax*; Intron absent, lanes (6) *Hemerocallis littorea*, (7) *Simethis mattiazii*, (8) *Caesia contorta*, (9) *Johnsonia pubescens*, (10) *Aloe purpurea*, (11) *Asphodelus albus*.

Figure 2.2

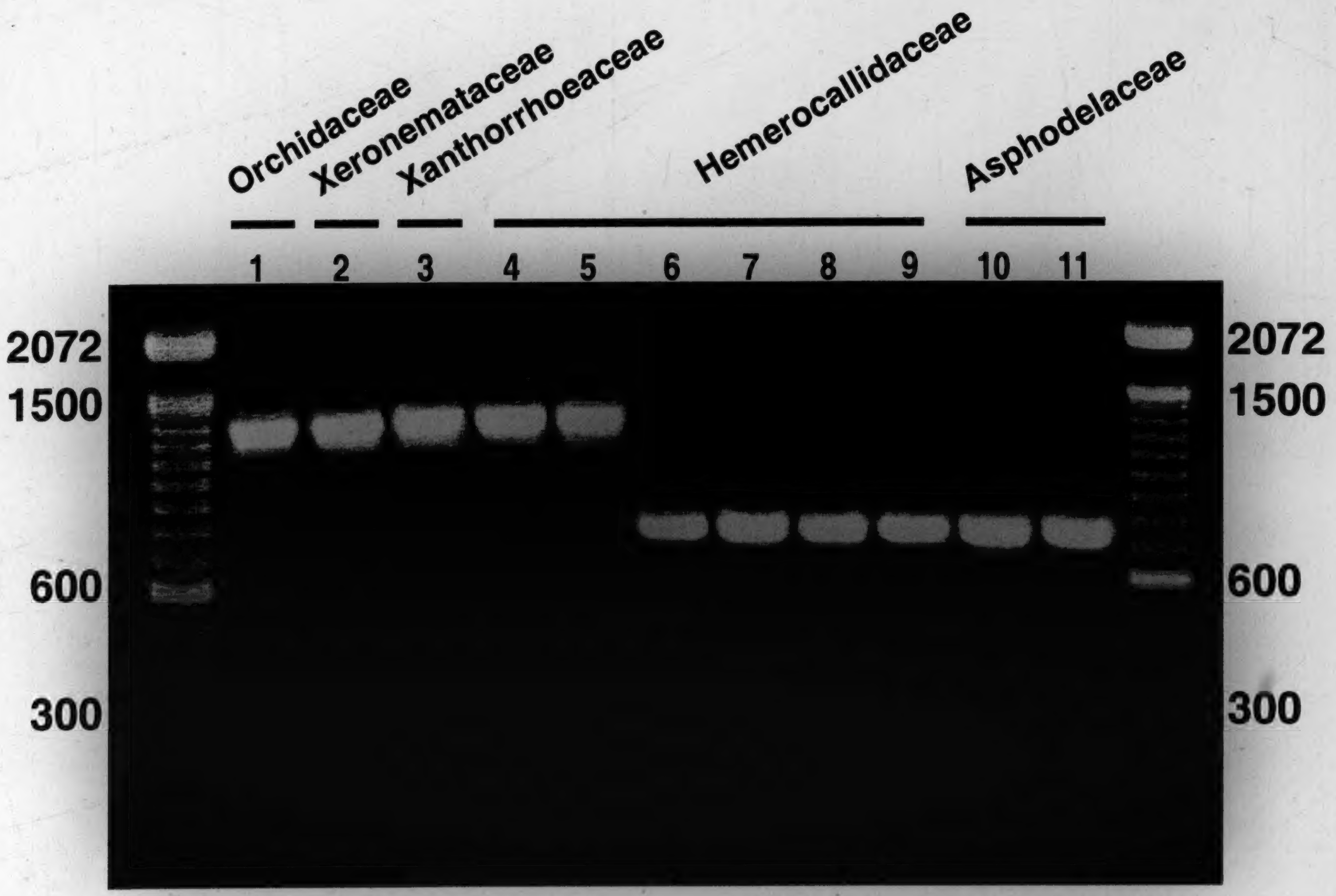

Orchidaceae
Xeronemataceae
Xanthorrhoeaceae
Hemerocallidaceae
Asphodelaceae
1
2
3
4
5
6
7
8
9
10
11
2072
1500
600
300
2072
1500
600
300

FIG. 2.3. Portion of one of four most-parsimonious trees of Asparagales inferred from a combined analysis of several chloroplast regions (*atp*B, *rbc*L and *trn*L-F; tree length = 5020 steps; CI = 0.426, RI = 0.541). Arrows indicate branches not found in the strict consensus of the most parsimonious trees. Branch lengths are calculated using ACCTRAN optimization; the scale bar indicates 50 changes. Bootstrap support values are indicated beside branches. The data set is based on the matrix of Fay *et al.* (2000), with four additional taxa (*Corynotheca*, *Pasithea*, *Stypandra* and *Tricoryne*; see text). Family-level circumscriptions follow APG (1998); the asterisk indicates tribe Johnsonieae Benth. (=Johnsoniaceae J.P. Lotsy). Outgroups (representatives of Commelinales, Liliales, Pandanales and Zingiberales) were included in the analysis, but are excluded here for clarity (see Fay *et al.* 2000 for details of outgroup relationships). Taxa in Asparagales that were sampled for 3'-*rps*12 intron status (for the same species or for congeners; see text), are indicated by filled and unfilled boxes, representing intron presence and absence, respectively. See Table 2.1 and the Appendix for a more complete survey of intron status.

Figure 2.3

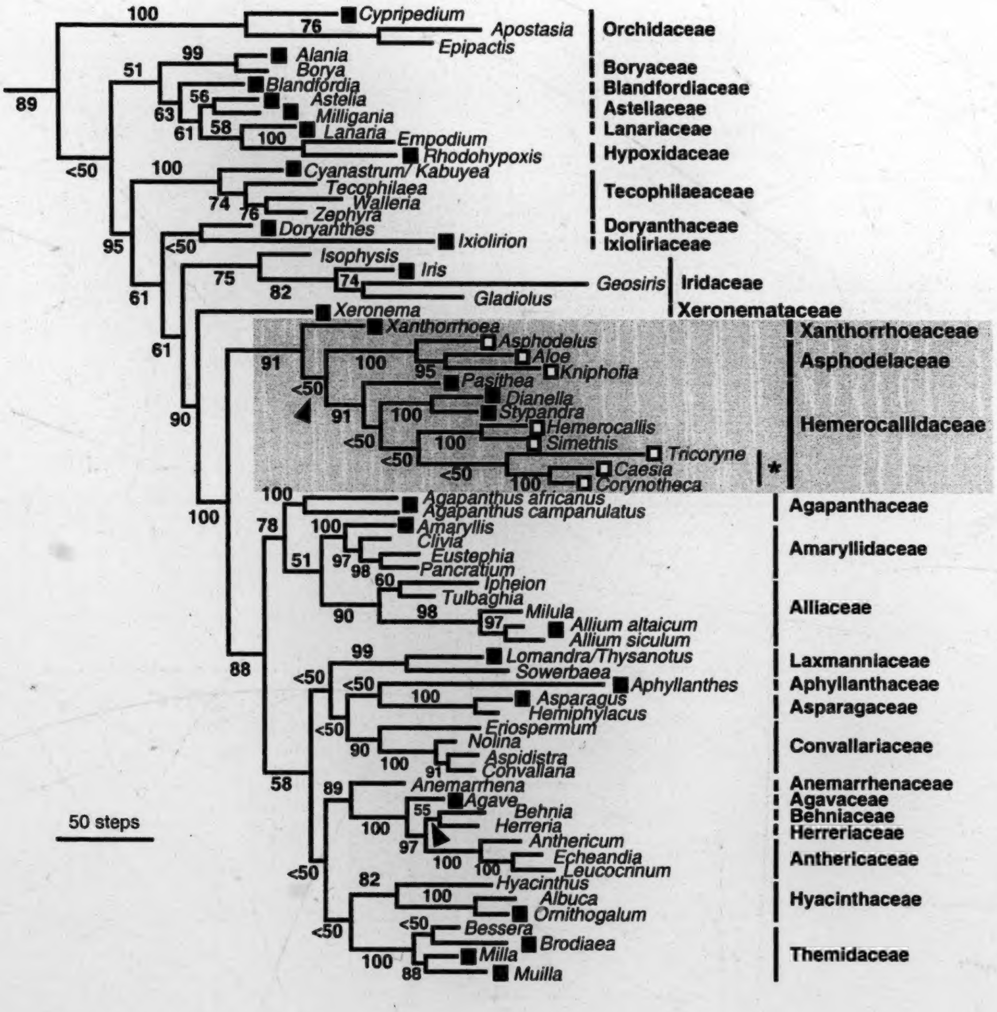

100
Cypripedium
76
Apostasia
Epipactis
Orchidaceae
89
99
Alania
51
Borya
Boryaceae
Blandfordia
Blandfordiaceae
56
Astelia
Asteliaceae
63
Milligania
58
61
Lanaria
Lanariaceae
100
Empodium
Rhodohypoxis
Hypoxidaceae
<50
100
Cyanastrum/ Kabuyea
Tecophilaea
Tecophilaeaceae
74
Walleria
76
Zephyra
Doryanthes
Doryanthaceae
95
<50
Ixiolirion
Ixioliriaceae
Isophysis
Iris
75
74
82
Geosiris
Iridaceae
Gladiolus
61
Xeronema
Xeronemataceae
61
Xanthorrhoea
Xanthorrhoeaceae
91
Asphodelus
100
Aloe
Asphodelaceae
95
Kniphofia
<50
Pasithea
Dianella
91
100
Stypandra
Hemerocallidaceae
90
Hemerocallis
<50
100
Simethis
<50
Tricoryne
*
<50
Caesia
100
Corynotheca
100
Agapanthus africanus
100
Agapanthus campanulatus
Agapanthaceae
78
100
Amaryllis
Clivia
Amaryllidaceae
51
97
Eustephia
98
Pancratium
60
Ipheion
Tulbaghia
Alliaceae
90
98
Milula
97
Allium altaicum
Allium siculum
99
Lomandra/Thysanotus
Laxmanniaceae
88
<50
Sowerbaea
Aphyllanthes
Aphyllanthaceae
<50
100
Asparagus
Asparagaceae
Hemiphylacus
<50
Eriospermum
90
Nolina
Convallariaceae
100
Aspidistra
91
Convallaria
58
89
Anemarrhena
Anemarrhenaceae
Agave
Agavaceae
55
Behnia
Behniaceae
50 steps
100
Herreria
Herreriaceae
97
Anthericum
100
100
Echeandia
Anthericaceae
Leucocrinum
<50
82
Hyacinthus
100
Albuca
Hyacinthaceae
Ornithogalum
<50
<50
Bessera
Brodiaea
100
Milla
Themidaceae
88
Muilla

FIG. 2.4. Flat-weighted parsimony reconstructions of the evolution of 3'-*rps*12 intron status in a subset of taxa from Asphodelaceae, Hemerocallidaceae, Xanthorrhoeaceae and relatives, for: (a) the shortest ML tree; (b) the shortest ML tree found using a backbone constraint that restricts intronless taxa to a monophyletic group (the constrained node is indicated with a hollow arrow); (c) suboptimal variants of the shortest ML tree (Fig. 2.4a) with *Hemerocallis* and *Tricoryne* forced into three different arrangements at the base of Hemerocallidaceae (the rearrangements are highlighted with arrows; see also Fig. 2.5). The optimal tree length, or the increase (δ) in tree score (-ln L) beyond the best tree, is indicated to the left of each tree. The results of SOWH-tests that assess whether the suboptimal trees are significantly longer than the best tree are also indicated in parentheses. The numbers and types of intron changes reconstructed on these trees are indicated in small boxes to the right of each tree.

Figure 2.4

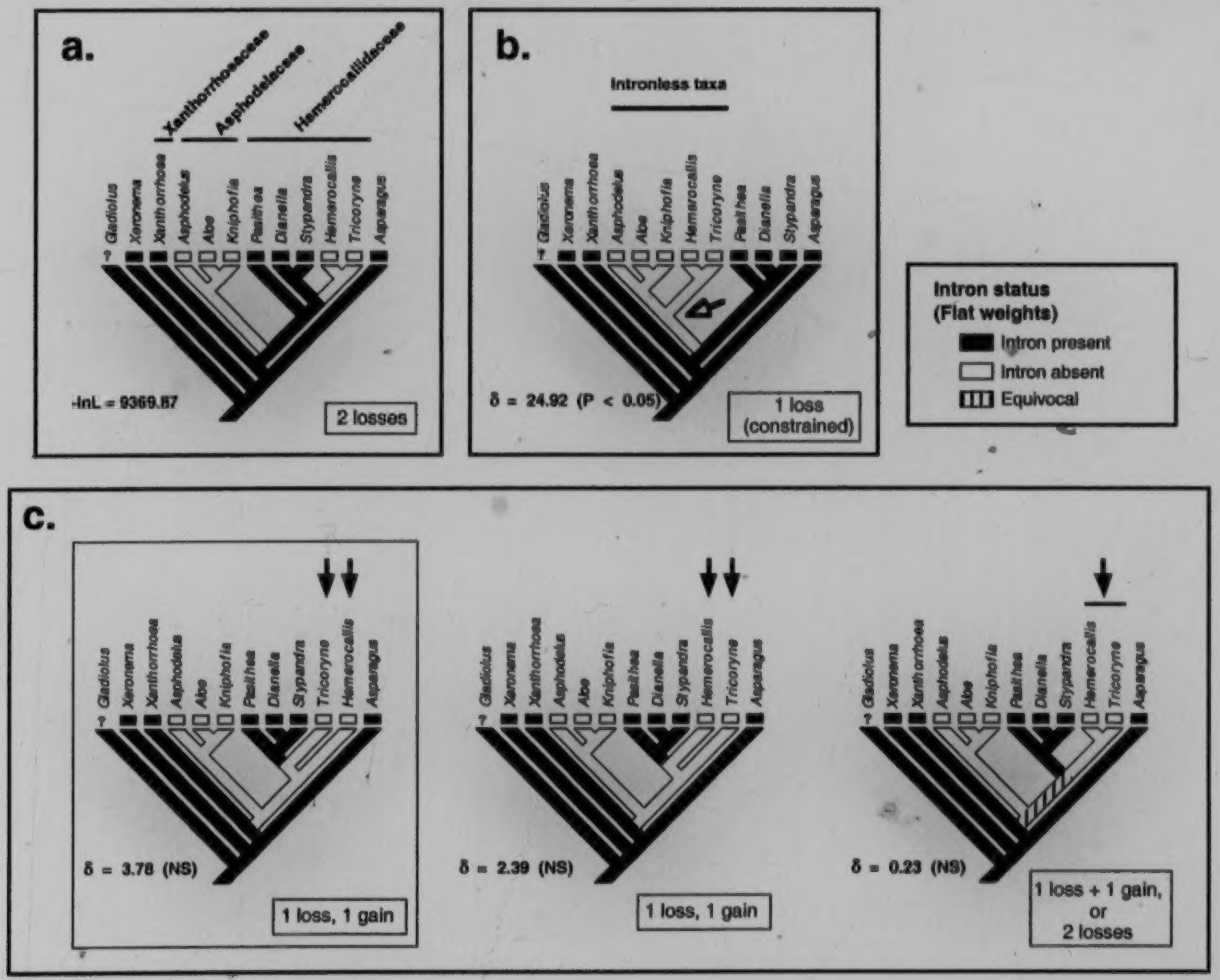
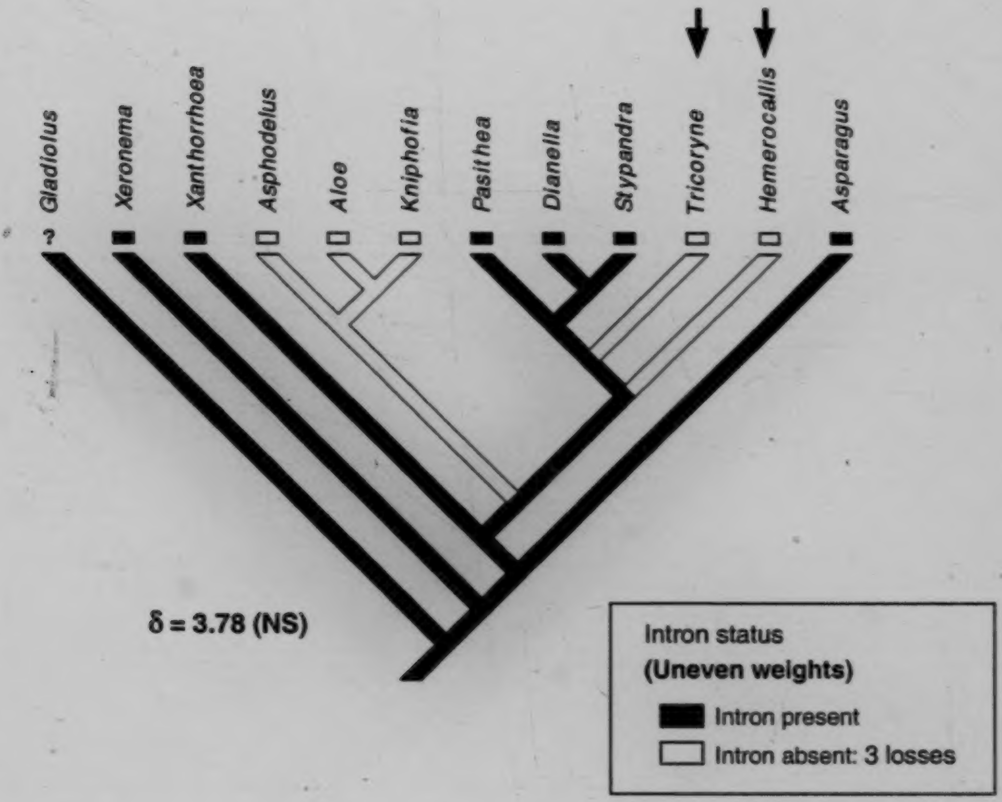

a.
Xanthorrhoeaceae
Asphodelaceae
Hemerocallidaceae
Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Pasithea
Dianella
Stypandra
Hemerocallis
Tricoryne
Asparagus
-lnL = 9369.87
2 losses
b.
Intronless taxa
Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Hemerocallis
Tricoryne
Pasithea
Dianella
Stypandra
Asparagus
δ = 24.92 (P < 0.05)
1 loss (constrained)
Intron status (Flat weights)
Intron present
Intron absent
Equivocal
c.
Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Pasithea
Dianella
Stypandra
Tricoryne
Hemerocallis
Asparagus
δ = 3.78 (NS)
1 loss, 1 gain
Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Pasithea
Dianella
Stypandra
Hemerocallis
Tricoryne
Asparagus
δ = 2.39 (NS)
1 loss, 1 gain
Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Pasithea
Dianella
Stypandra
Hemerocallis
Tricoryne
Asparagus
δ = 0.23 (NS)
1 loss + 1 gain, or 2 losses

FIG. 2.5. Parallel intron loss reconstructed using an uneven parsimony weighting scheme on a suboptimal tree (Fig. 2.4c, left-hand tree). This tree is not significantly longer than the optimal ML tree (Fig. 2.4a) from which it was derived (rearrangements highlighted with arrows), according to an SOWH-test. The increase (δ) in tree score (-ln L) beyond the best tree is indicated. The reconstruction shown here uses a parsimony-based step matrix that disfavors 3'-*rps*12 intron gain relative to its loss by a 2.025: 1 margin. All weighting schemes assessed for this tree that employ at least this much bias yield the same reconstruction, with three intron losses inferred among several closely related lineages of Asphodelaceae and Hemerocallidaceae.

Figure 2.5


Gladiolus
Xeronema
Xanthorrhoea
Asphodelus
Aloe
Kniphofia
Pasithea
Dianella
Stypandra
Tricoryne
Hemerocallis
Asparagus
?
δ = 3.78 (NS)
Intron status
(Uneven weights)
■ Intron present
□ Intron absent: 3 losses

Appendix 2.1. Status of 3'-*rps*12 intron in selected taxa outside Asparagales.

Taxon	Source	Intron
Monocots		
Acorus calamus L. (Acoraceae; Acorales)	Denver Botanic Garden, CO, no voucher (RGO 97-149 DNA)	Present [1]
Acorus gramineus Aiton (Acoraceae; Acorales)	Rothwell and Williams s.n. (ALTA)	Present [1]
Ananas comosus (L.) Merr. (Bromeliaceae; Poales)	H. S. Rai 1003 (ALTA)	Present [1]
Anticlea elegans (Pursh) Rydberg (Melanthiaceae; Liliiales)	M. A. McPherson 68-990704 (ALTA)	Present [1]
Anigozanthos flavidus D. C. (Haemodoraceae; Commelinales)	R. Neyland 1884 (MCN)	Present [1]
Burmannia capitata Mart. (Burmanniaceae; Dioscoreales)	R. Neyland 958 (MCN)	Present [1]
Campynema lineare Labill. (Campynemataceae; Liliales)	Walsh 3488 (MEL)	Present
Canna edulis Ker Gawl. (Cannaceae; Zingiberales)	M. A. McPherson 020425-1 (ALTA) [2]	Present
Cartonema philydroides Muell. (Commelinaceae; Commelinales)	T. Evans s.n.	Present [1]
Dasypogon hookeri J. R. Drumm. (Dasypogonaceae; Commelinoids)	M. W. Chase 430 (NCU)	Present [1]
Dioscorea bulbifera L. (Dioscoreaceae; Dioscoreales)	EPO Biology, U. Colorado, Boulder, CO,	Present [1]

	no voucher (RGO 97-151 DNA)	
Eichhornia crassipes (Mart.) Solms (Pontederiaceae; Commelinales)	Barrett 814 (TRT)	Present
Ensete ventricosum (Welw.) Cheesman (Musaceae; Zingiberales)	Kress 96-5372 (US)	Present [1]
Eriocaulon compressum Lam. (Eriocaulaceae; Poales)	M. M. Unwin #241 (MU)	Present [1]
Hanguana malayana (Jack) Merr. (Hanguanaceae, Commelinales)	Sirirugsa s.n.	Present [1]
Helmholtzia glaberrima (Hook. f.) Caruel (Philydraceae; Commelinales)	No voucher (Graham DNA #02091)	Present [1]
Heteranthera limosa (Swartz) Willd. (Pontederiaceae; Commelinales)	Barrett 1054 (TRT)	Present
Hydrothrix gardneri Hook. f. (Pontederiaceae; Commelinales)	Barrett 1414 (TRT)	Present [1]
Joinvillea plicata (Hook. f.) Newell & Stone (Joinvilleaceae; Poales)	Thein 84 (NO)	Present [1]
Leucojum aestivum L. (Amaryllidaceae; Asparagales)	SW Graham 00-4-2 ALTA	Present
Lilium superbum L. (Liliaceae; Liliales)	M. W. Chase 112 (NCU)	Present [1]
Maranta leuconeura E. Morren (Marantaceae; Zingiberales)	M. A. McPherson 020425-2 (ALTA) [2]	Present
Mayaca fluviatilis Aubl. (Mayacaceae; Poales)	J. R. Abbott s.n.	Present [1]
Monochoria korsakowii Reg. & Maack (Pontederiaceae; Commelinales)	Barrett 1415 (TRT)	Present
Oryza sativa L. (Poaceae; Poales)	GenBank plastid genome: NC_001320.1	Present [1]
Palisota bogneri J. P. M. Brenan (Commelinaceae; Commelinales)	Rothwell & Stockey 90 (ALTA)	Present [1]

Philydrum lanuginosum Banks and Sol. ex Gaertn. (Philydraceae; Commelinales)	Graham & Barrett 1 (TRT)	Present[1]
Roystonea princeps (Becc.) Burret (Arecaceae; Arecales)	E. Santiago #J-4, (UPR)	Present[1]
Sagittaria latifolia Willd. (Alismataceae; Alismatales)	Barrett s.n. (TRT)	Present[1]
Scheuchzeria palustris L. (Scheuchzeriaceae; Alismatales)	M. Waterway & S.W. Graham 97-60 (ALTA)	Present[1]
Spathiphyllum wallisii Hort. (Araceae; Arecales)	M. W. Chase 210 (NCU)	Present[1]
Stemona tuberosa Lour. (Stemonaceae; Pandanales)	Rothwell and Stockey 46 (ALTA)	Present[1]
Strelitzia reginae Aiton (Strelitziaceae; Zingiberales)	S. W. Graham s.n. (ALTA)	Present[1]
Talbotia elegans Balf. (Velloziaceae; Pandanales)	Rothwell & Stockey 48 (ALTA)	Present[1]
Tofieldia glutinosa (Michx.) Pers. (Tofieldiaceae; Alismatales)	S. Stefanovic s.n. (ALTA)	Present[1]
Triticum aestivum L. (Poaceae; Poales)	GenBank plastid genome: NC_002762.1	Present[1]
Typha angustifolia L. (Typhaceae; Typhales)	Graham 1040 (TRT)	Present[1]
Typha latifolia L. (Typhaceae; Typhales)	M. A. McPherson 010819-3 (ALTA)	Present[1]
Xiphidium caeruleum Aubl. (Haemodoraceae; Commelinales)	SDSU greenhouse (coll. M. Simpson) (SWG 5.7.94 DNA)	Present[1]
Xyris jupicai Rich. (Xyridaceae; Poales)	D. Goldman 1766 (BH)	Present[1]

Zea mays L. (Poaceae; Poales)	GenBank plastid genome: NC_001666.2	Present [1]
Other angiosperms		
Amborella trichopoda Baill. (Amborellaceae; Amborellales)	M. P. Simmons 1846 (GH)	Present [1]
Arabidopsis thaliana (L.) Heynh. (Brassicaceae; Brassicales)	GenBank plastid genome: NC_000932.1	Present [1]
Asarum canadense L. (Aristolochiaceae; Piperales)	Ann Arbor, MI, no voucher (RGO 748 DNA)	Present [1]
Cabomba caroliniana A. Gray (Nymphaeaceae; Nymphaeales)	Les s.n. (CONN)	Present [1]
Calycanthus floridus L. (Calycanthaceae; Laurales)	Matthaei Bot. Gard., Ann Arbor, MI, no voucher (RGO 307 DNA)	Present [1]
Ceratophyllum demersum L. (Ceratophyllaceae; Ceratophyllales)	Les s.n. (CONN)	Present [1]
Cercidiphyllum japonicum Sieb. & Zucc. (Cercidiphyllaceae; Saxifragales)	RGO 90-016 (WTU)	Present [1]
Drimys winteri J. R. Forst. And G. Forst. (Winteraceae; Winterales)	RGO 97-13 (WTU)	Present [1]
Epifagus virginiana (L.) W. Bart. (Orobanchaceae; Lamiales)	GenBank plastid genome: NC_001568.1	Present [1]
Glycine max (L.) Merr. (Fabaceae; Fabales)	GenBank accession: X07675	Present [1]
Illicium parviflorum Michx. ex Vent (Illiciaceae; Austrobaileyales)	Naczi 2784 (MICH)	Present [1]
Lactoris fernandeziana Phil. (Lactoridaceae; Piperales)	Stuessy et al. 11591 (OS)	Present [1]

Liriodendron tulipifera L. (Magnoliaceae; Magnoliales)	RGO 97-9 (WTU)	Present [1]
Lotus japonicus (Regel) K. Larsen (Fabaceae; Fabales)	GenBank complete plastid genome: NC_002694. 1	Present [1]
Medicago truncatula Gaertn. (Fabaceae; Fabales)	GenBank complete plastid genome: NC_003119.6	Absent [1]
Nicotiana tabacum L. (Solanaceae; Solanales)	GenBank accession: Z00044	Present [1]
Nymphaea odorata Aiton (Nymphaeaceae; Nymphaeales)	Les s.n. (CONN)	Present [1]
Oenothera elata Humboldt, Bonpland and Kunth (Onagraceae: Myrtales)	GenBank complete plastid genome: NC_002693.1	Present [1]
Pisum sativum ~~L.~~ (Fabaceae; Fabales)	R. G. Olmstead 98-60 (WTU)	Absent [1]
Saururus cernuus L. (Saururaceae; Piperales)	RGO 88-006 (WTU)	Present [1]
Spinacia oleracea L. (Amaranthaceae; Caryophyllales)	GenBank plastid genome: NC_002202.1	Present [1]
Trochodendron aralioides P. F. Siebold & J. G. Zuccarini (Trochodendraceae; Trochodendrales)	RGO 97-8 (WTU)	Present [1]
Other land plants		
Ginkgo biloba L. (Ginkgoaceae; Ginkgoales)	SWG 97-IV (1) (WTU)	Present [1]
Gnetum gnemon L. (Gnetaceae; Gnetales)	RGO 97-16 (WTU)	Present [1]
Marchantia polymorpha L. (Marchantiaceae; Marchantiales)	GenBank plastid genome: NC_001319.1	Present [1]

Pinus thunbergii Parl. (Pinaceae; Coniferales)	GenBank plastid genome: NC_001631.1	Present [1]
Psilotum nudum (L.) P. Beauv. (Psilotaceae; Psilotales)	GenBank plastid genome: NC_003386.1	Present [1]

---

[1] Intron status confirmed by DNA sequencing or by examination of GenBank accession.

[2] Cultivar.

Bibliography

ANGIOSPERM PHYLOGENY GROUP. 1998. An ordinal classification for the families of flowering plants. Annals Missouri Botanical Garden 85: 531-553.

BARRY, D. and J. A. HARTIGAN. 1987. Asynchronous distance between homologous DNA sequences. Biometrics 43: 261-276.

BOGLER, D. J. 1998. Nolinaceae. Pp. 392-397 in *The Families and Genera of Vascular Plants. Volume III Flowering Plants. Monocotyledons: Lilianae (except Orchidaceae)*, eds. K. Kubitzki. Berlin: Springer.

BONEN, L. and J. VOGEL. 2001. The ins and outs of group II introns. Trends in Genetics 17: 322-331.

CAMPAGNA, M. L. and S. R. DOWNIE. 1998. The intron in chloroplast gene *rpl*16 is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. Transactions of the Illinois State Academy of Science 91: 1-11.

CHASE, M. W., A. Y. DE BRUIJN, A. V. COX, G. REEVES, P. J. RUDALL, M. A. T. JOHNSON and L. E. EGUIARTE. 2000a. Phylogenetics of Asphodelaceae (Asparagales): an analysis of plastid *rbcL* and *trnL-F* DNA sequences. Annals of Botany 86: 935-951.

CHASE, M. W., M. R. DUVALL, H. G. HILLS, J. G. CONRAN, A. V. COX, L. E. EGUIARTE, J. HARTWELL, M. F. FAY, L. R. CADDICK, K. M. CAMERON, and S. HOOT. 1995.

Molecular phylogenetics of Lilianae. Pp. 109-137 in *Monocotyledons: systematics and evolution*, eds. P. J. Rudall, P. J. Cribb, D. F. Cutler, C. J. Humphries. Kew: Royal Botanical Gardens.

CHASE, M. W., P. J. RUDALL and J. G. CONRAN. 1996. New circumscriptions and a new family of asparagoid lilies: genera formerly included in Anthericaceae. Kew Bulletin 51: 667-680.

CHASE, M. W., D. E. SOLTIS, P. S. SOLTIS, P. J. RUDALL, M. F. FAY, W. H. HAHN, S. SULLIVAN, J. JOSEPH, M. MOLVRAY, P. J. KORES, T. J. GIVNISH, K. J. SYTSMA and J. C. PIRES. 2000b. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. Pp. 3-16 in *Monocots: systematics and evolution*, eds. K. L. Wilson and D. A. Morrison. Victoria: CSIRO Publishing, Collingwood.

CHO, Y., Y. -L. QIU, P. KUHLMAN and J. D. PALMER. 1998. Explosive invasion of plant mitochondria by a group I intron. Proceedings of the National Academy of Sciences of the United States 95: 14244-14249.

CLEGG, M. T., B. S. GAUT, G. H. JR. LEARN and B. M. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. Proceedings of the National Academy of Sciences of the United States 91: 6795-6801.

CLIFFORD, H. T. and J. G. CONRAN. 1998. Johnsoniaceae. Pp. 336-340 in *The Families and Genera of Vascular Plants Volume III Flowering Plants Monocotyledons: Lilianae (except Orchidaceae)*, eds. K. Kubitzki. Berlin: Springer.

CLIFFORD, H. T., R. J. F. HENDERSON and J. G. CONRAN. 1998. Hemerocallidaceae. Pp. 245-252 in *The Families and Genera of Vascular Plants Volume III Flowering Plants Monocotyledons: Lilianae (except Orchidaceae)*, eds. K. Kubitzki. Berlin: Springer.

DAI, L. and S. ZIMMERLY. 2002. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucleic Acid Research 30: 1091-1102.

DAVIS, P. H. and V. H. HEYWOOD. 1973. Principles of Angiosperm Taxonomy. Krieger, Huntington.

DONOGHUE, M. J. and D. D. ACKERLY. 1996. Phylogenetic uncertainties and sensitivity analyses in comparative biology. Philosophical Transactions Royal Society of London B 351: 1241-1249.

DOWNIE, S. R. and J. D. PALMER. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. Pp. 14-35 in *Molecular systematics of plants*, eds. P. S. Soltis, D. E. Soltis and J. J. Doyle. New York: Chapman and Hall Inc.

DOWNIE, S. R., D. S. KATZ-DOWNIE, E. J. ROGERS, H. L. ZUJEWSKI and E. SMALL. 1998. Multiple independent losses of the plastid *rpo*C1 intron in *Medicago*

(Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. Canadian Journal of Botany 76: 791-803.

DOWNIE, S. R., E. LLANAS and D. KATZ-DOWNIE. 1996. Multiple independent losses of the *rpo*C1 intron in angiosperm chloroplast DNA's. Systematic Botany 21: 135-151.

DOWNIE, S. R., R. G. OLMSTEAD, G. ZURAWSKI, D. E. SOLTIS, P. S. SOLTIS, J. C. WATSON and J. D. PALMER. 1991. Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: molecular and phylogenetic implications. Evolution 45: 1245-1259.

DOYLE, J. J. and J. L. DOYLE. 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.

DOYLE, J. J., J. L. DOYLE and J. D. PALMER. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Systematic Botany 20: 272-294.

DRESSLER, R. L. 1993. *Phylogeny and classification of the orchid family*. Dioscorides Press, Portland, Oregon.

FARRIS, J. S. 1970. Methods for computing Wagner trees. Systematic Zoology 19: 83-92.

FAY, M. F., P. J. RUDALL, S. SULLIVAN, K. L. STOBART, A. Y. DE BRUIJN, G. REEVES, F. QAMARUZ-ZAMAN, W. -P. HONG, J. JOSEPH, W. J. HAHN, J. G. CONRAN and M. W.

CHASE. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. Pp. 360-371 in *Monocots: systematics and evolution*, eds. K. L. Wilson and D. A. Morrison. Victoria: CSIRO Publishing, Collingwood.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17: 368-376.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

FITCH, W. M. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Systematic Zoology 20: 406-416.

—–. 1984. Cladistic and other methods: problems, pitfalls, and potentials. Pp. 221-254 *in* T. Duncan and T. F. Steussy, eds. Perspectives on the reconstruction of evolutionary history. Columbia University Press, New York.

FREYER, R., K. NECKERMANN, R. M. MAIER and H. KOSSEL. 1995. Structural and functional analysis of plastid genomes from parasitic plants: loss of an intron within the genus *Cuscuta*. Current Genetics 27: 580-586.

GOLDMAN, N., J. P. ANDERSON and A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. Systematic Biology 49: 652-670.

GOREMYKIN, V., V. BOBROVA, J. PAHNKE, A. TROITSKY, A. ANTONOV and W. MARTIN. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support gnetalean affinities of angiosperms. Molecular Biology and Evolution 13: 383-396.

GRAHAM, S. W. and R. G. OLMSTEAD. 2000a. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany 87: 1712-1730.

—. 2000b. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. Current Genet.ics 37: 183-188.

GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS and R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int. J. Plant Sci. 161: S83-S96.

HARTIGAN, J. A. 1973. Minimum mutation fits to a given tree. Biometrics 29: 53-65.

HASEGAWA, M., K. KISHNINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 32: 443-445.

HILDEBRAND, M., R. B. HALLICK, C. W. PASSAVANT and D. P. BOURQUE. 1988. Trans-splicing in chloroplasts: The *rps12* loci of *Nicotiana tabacum*. Proceedings of the National Academy of Sciences of the United States 85: 372-376.

HOOT, S. B., and J. D. PALMER. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. Journal of Molecular Evolution 38: 274-281.

HOOT, S. B., A. CULHAM and P. R. CRANE. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in Lardizabalaceae. Annals of the Missouri Botanical Garden 82: 194-207.

HUELSENBECK, J. P. and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and Systematics 28: 437-466.

HUELSENBECK, J. P., L. LARGET, R. E. MILLER and F. RONQUIST. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology 51: 673-688.

JOLY, S., L. BROUILLET and A. BRUNEAU. 2001. Phylogenetic implications of the multiple losses of the mitochondrial *cox*II.*i*3 intron in the angiosperms. International Journal of Plant Sciences 162: 359-373.

JUKES, T. H. and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21-132 in *Mammalian Protein Metabolism* , eds. H. N. Munro. Academic Press, New York.

KAJITA, T., H. OHASHI, Y. TATEISHI, S. D. BAILEY and J. J. DOYLE. 2001. *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. Systematic Botany 26: 515-536.

KOLLER, B., H. FROMM, E. GALUN and M. EDELMAN. 1987. Evidence for in vivo *trans* splicing of pre-mRNAs in tobacco chloroplasts. Cell 48: 111-119.

KOHN, J. R., S. W. GRAHAM, B. MORTON, J. J. DOYLE and S. C. H. BARRETT. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. Evolution 50: 1454-1469.

LAI, M., J. SCEPPA, J. A. BALLENGER and J. J. DOYLE. 1997. Polymorphism for the presence of the *rpl*2 intron in chloroplast genomes of *Bauhinia* (Leguminosae). Syst. Bot. 22: 519-528.

LANAVE, C., G. PREPARATA, C. SACCONE and G. SERIO. 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20: 86-93.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological data. Systematic Biology 50: 913-925.

MADDISON, W. P. and D. R. MADDISON. 1992. *MacClade: analysis of phylogeny and character evolution, Version 3.0, Computer program and documentation.* Sunderland: Sinauer Associates Inc.

—. 2001. *MacClade: analysis of phylogeny and character evolution, Version 4.03, Computer program and documentation.* Sunderland: Sinauer Associates Inc.

MARTIN, W., B. STOEBE, V. GOREMYKIN, S. HANSMANN, M. HASEGAWA and K. V. KOWALLIK. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393: 162-165.

MOOERS, A. Ø. and D. SCHLUTER. 1999. Reconstructing ancestor states with maximum likelihood: support for one- and two-rate models. Systematic Biology 48: 623-633.

MUASYA, A. M., D. A. SIMPSON, A. CULHAM and M. W. CHASE. 1998. An assessment of suprageneric phylogeny in Cyperaceae using *rbcL* DNA sequences. Plant Systematics and Evolution 211: 257-271.

NICKRENT, D. L., R. J. DUFF, A. E. COLWELL, A. D. WOLFE, N. D. YOUNG, K. E. STEINER and C. W. DEPAMPHILIS. 1999. Molecular phylogenetics and evolutionary studies of parasitic plants. Pp. 211-241 in *Molecular Systematics of Plants II: DNA Sequencing*, eds. D. E. Soltis, P. S. Soltis and J. J. Boston: Kluwer Academic Publishers.

OMLAND, K. E. 1997. Examining two standard assumptions of ancestral reconstructions: repeated loss of dimorphism in dabbling ducks (Anatini). Evolution 51: 1636-1646.

——. 1999. The assumptions and challenges of ancestral state reconstructions. Systematic Biology 48: 604-611.

PAGEL, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society of London B 255: 37-45.

——. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta 26: 331-348.

PALMER, J. D. 1991. Plastid chromosomes: structure and evolution. Pp. 5-53 in *Volume 7A, Cell culture and somatic cell genetics of plants*, eds. L. Bogorad and I. K. Vasil. San Diego: Academic Press.

PALMER, J. D., B. OSORIO, J. ALDRICH and W. F. THOMPSON. 1987. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Current Genetics 11: 275-286.

PALMER, J. D., B. OSORIO and W. F. THOMPSON. 1988. Evolutionary significance of inversions in legume chloroplast DNAs. Current Genetics 14: 65-74.

RAMBAUT, A. and N. C. GRASSLY. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13: 235-238.

RODRÍGUEZ, F., J. L. OLIVER, A. MARÍN and J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142: 485-501.

RUDALL, P. J., C. A. FURNESS, M. W. CHASE and M. F. FAY. 1997. Microsporogenesis and pollen sulcus type in Asparagales (Lilianae). Canadian Jounal of Botany 75: 408-430.

SANDERSON, M. J. 1993. Reversibility in evolution: a maximum likelihood approach to character gain-loss bias in phylogenies. Evolution 47: 236-252.

SMITH, G. F. and B.–E. VAN WYK. 1998. Asphodelaceae. Pp. 130-140 in *The Families and Genera of Vascular Plants. Volume III Flowering Plants. Monocotyledons: Lilianae (except Orchidaceae)*, eds. K. Kubitzki. Berlin: Springer.

SWOFFORD, D. L. 2002. "*PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*," Computer program and documentation. Sinauer, Sunderland, MA.

SWOFFORD, D. L. and W. P. MADDISON. 1987. Reconstructing ancestral character states under Wagner parsimony. Mathematical Biosciences 87: 199-229.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL and D. M. HILLIS. 1996. Phylogenetic Inference. Pp. 407-543 in *Molecular Systematics, Second edition*, eds. D. M. Hillis, C. Moritz and B. K. Mable. Sunderland: Sinauer.

TABERLET, P., L. GIELLY, G. PAUTOU and J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.

TAVARÉ, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences. 17: 57-86.

WOLFE, K. H, W. –H. LI and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States 84: 9054-9058.

ZAITA, N., K. TORAZAWA, K. SHINOZAKI and M. SUGIURA. 1987. *Trans*-splicing in vivo: joining of transcripts from the 'divided' gene for ribosomal protein S12 in the chloroplasts of tobacco. Federation of European Biochemical Societies 210: 153-156.

ZURAWSKI, G., M. T. CLEGG and A. H. D. BROWN. 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106: 735-749.

~~ CHAPTER THREE ~~

PHYLOGENETIC RELATIONSHIPS AMONG THE MAJOR LINEAGES OF ASPARAGALES BASED ON A LARGE PLASTID DATA SET, WITH A SPECIAL FOCUS ON THE POSITION OF APHYLLANTHACEAE

Introduction

As circumscribed in the recent classification by the Angiosperm Phylogeny Group (APG, 1998), the order Asparagales is one of the most species-rich groups of flowering plants. This cosmopolitan order includes many taxa that are economically important or aesthetically pleasing to humans, including the daffodils and allies in the Amaryllidaceae (ca. 870 species), irises and relatives (Iridaceae; ca. 1,750 species), onions and relatives (Alliaceae; ca. 645 species), and the orchids (some 19,500 recognized species; Judd *et al.*, 2002). The latter family contributes a substantial fraction of extant angiosperm species diversity. As a consequence, Asparagales comprise ca. 27,000 of the 65,000 currently recognized species of monocots, or about a tenth of the total number of flowering plant species (Judd *et al.*, 2002). The order encompasses a tremendous range of floral diversity, and a broad variety of life forms and habits (e.g., Kubitzki, 1998a). This morphological diversity is reflected in recent taxonomic schemes, which recognize numerous families in Asparagales (30 in Dahlgren *et al.*, 1985, 29 in APG, 1998). Much recent systematic work (reviewed briefly below) has focussed on clarifying the taxonomic circumscription of the order and its constituent families.

The order in its modern sense dates back to Huber (1969), who defined it based primarily on seed characteristics. Most families in Asparagales have a completely collapsed inner seed coat, and an outer epidermis that is usually either obliterated (in fleshy fruited taxa), or present with a carbonaceous phytomelan crust (in dry-fruited taxa). Huber's taxonomic treatment was refined by Dahlgren and colleagues (Dahlgren, 1980; Dahlgren and Clifford, 1982; Dahlgren and Rasmussen, 1983; Dahlgren *et al.*, 1985), and substantiated and further clarified using molecular and other lines of evidence (e.g., APG, 1998). While not completely congruent with each other (for example, APG recognize the inclusion of Orchidaceae and Iridaceae in the order, the others do not), these schemes differ substantially from other modern taxonomic treatments, typified at one extreme by the views of Cronquist (1981, 1988).

Cronquist's taxonomic scheme for the monocots departs strongly from modern phylogenetic studies, most spectacularly concerning his treatment of the "liliaceous" monocots (reviewed in Dahlgren and Clifford, 1982). His system is still taught in basic plant systematics courses, and is widely used in recent textbooks and floras (e.g., the ongoing series of the Flora of North America; Oxford University Press). Cronquist disregarded the seed anatomy evidence supporting Asparagales, and lumped many of its component taxa into a single large family, Liliaceae (following Krause, 1930), in the order Liliales (Chapter 1, Table 1.1). The family Liliaceae *sensu* Cronquist encompasses taxa that possess conspicuous, perfect flowers, typically with six stamens, with the ovary usually superior, and with raphides present in a subset of cells. He characterized members of Liliaceae as "…geophytes with soft annual (rarely perennial) leaves, the stems herbaceous or nearly so and usually without secondary growth" (Cronquist 1988: p. 494). Cronquist recognized that this family was diverse and poorly developed taxonomically, but he could not come up with a reasonable scheme for subdividing it. In contrast, Huber, and Dahlgren and colleagues, were taxonomic "splitters." They used rather narrow taxonomic concepts in Asparagales and the other petaloid monocots to avoid the recognition of this type of heterogeneous assemblage (Dahlgren and Clifford, 1982).

Recent systematic work on the monocots develop themes that were first suggested in pioneering molecular studies that used the conserved plastid (chloroplast) gene *rbc*L (Chase *et al.*, 1993, 1995; Duvall *et al.*, 1993) to address angiosperm and monocot phylogenetics on a broad scale. These studies recently culminated in the order-level treatment of the flowering plants by APG (1998), which was based primarily on analyses of DNA sequence data from three genes (two plastid genes, *atp*B, *rbc*L, and the nuclear small subunit ribosomal DNA locus, or "18S rDNA"). More recent molecular studies have focussed on the monocots as a whole (Chase *et al.*, 2000a) and Asparagales in particular (Fay *et al.*, 2000). All of these studies used one or a few genes sampled for a relatively small number of "exemplar" taxa that were chosen to broadly represent taxonomic diversity. Of the studies that most comprehensively represent Asparagales, Chase *et al.* (1995) sampled *rbc*L for 83 species in the order (and 89 other monocots; a data set later expanded by Rudall *et al.*, 1997, with the inclusion of 27 more Asparagales *rbc*L sequences), and Fay *et al.* (2000) sampled three noncontiguous plastid regions (*atp*B, *rbc*L, and *trn*L to *trn*F, including associated noncoding regions) for 64 species in Asparagales and 14 outgroup taxa.

These and other contemporary studies have greatly clarified our understanding of Asparagales systematics. With a few rather radical exceptions, much of this recent phylogenetic evidence supports the broad details of the taxonomic scheme of Dahlgren *et al.* (1985; reviewed in greater detail in Chapter 1). Asparagales *sensu* APG (1998) corresponds to a clade that includes Orchidaceae and Iridaceae, and that now excludes most or all of the taxa accepted by Dahlgren *et al.* (1985) in Calectasiaceae, Dasypogonaceae, Hanguanaceae, Luzuriagaceae and Philesiaceae. Several families recognized by Dahlgren *et al.* (1985) were synonymized with other Asparagales in APG 1998 (see Table 1.1, footnote 5). Each expanded family (Convallariaceae, Agavaceae and Hemerocallidaceae) corresponds to a clade of closely related and intermingled taxa in the *rbc*L-based study of Chase *et al.* (1995). Other small families (Anemarrhenaceae, Behniaceae, Boryaceae and Xeronemataceae) were erected (Chase *et al.*, 1996, 1997, 2000b; Conran *et al.*, 1997) to accommodate distinct phylogenetic lineages observed in molecular analyses. Agapanthaceae and Themidaceae were also resurrected to accommodate the nonmonophyly of several families as circumscribed by Dahlgren *et al.* (1985) (for Agapanthaceae: APG, 1998; Kubitzki, 1998b, and see Fay and Chase, 1996; for Themidaceae: Fay and Chase, 1996, and see Pires *et al.*, 2001; Pires and Sytsma, 2002). Both of these families consist of taxa previously considered to belong to Alliaceae.

Only a few families still recognized in Asparagales are altered radically from their corresponding taxonomic conceptions in Dahlgren *et al.* (1985). The two most obvious examples are Anthericaceae and Laxmanniaceae (= Lomandraceae). The circumscription of the former family in Dahlgren *et al.* (1985) yields one of the more grossly polyphyletic families with respect to analyses of *rbc*L sequence data (Chase *et al.*, 1995, 1996; Rudall *et al.*, 1997). As a consequence, Anthericaceae are now reduced to around eight genera (Conran, 1998), including *Anthericum* and *Chlorophytum* (spider plant), with its other taxa now placed in other families of Asparagales, including Boryaceae, Hemerocallidaceae, and Laxmanniaceae. Laxmanniaceae also now incorporate fragments of Dasypogonaceae and Luzuriagaceae (both *sensu* Dahlgren *et al.*, 1985) that were left behind in Asparagales after the removal of the latter two families to the commelinoid monocots and Liliales, respectively (Chase *et al.*, 1996).

Although family-level circumscriptions in the order are now fairly well understood (see Kubitzki, 1998a), substantial questions remain about higher-order relationships among the major lineages of Asparagales. The most comprehensive phylogenetic study of the order, that of Fay *et al.* (2000), left multiple aspects of the backbone of Asparagales relationship ambiguous or poorly supported (Chapter 2, Fig. 2.3). It is also suspected that long-branch attraction (Felsenstein, 1978; Hendy and Penny, 1985) may interfere with phylogenetic inference at some levels in the order. Fay *et al.* (2000) suggested that uncertain relationships among a cluster of closely related families (including Aphyllanthaceae, Asparagaceae, Convallariaceae, Hyacinthaceae and Laxmanniaceae) are a consequence of the long branch subtending the monotypic family Aphyllanthaceae.

These uncertainties frustrate attempts to reconstruct character evolution in the order. We can test hypotheses of evolution (e.g., Donoghue, 1989) and determine the identity of morphological characters (synapomorphies) that support major phylogenetic subdivisions (see Judd *et al.*, 2002, for examples in Asparagales) by performing character reconstructions on phylogenies, but ideally this requires that trees are resolved with a high degree of confidence. At present, it is not completely clear what the relationship is of Asparagales to the rest of the monocots, how broadly circumscribed the order should be, nor what the relationships are among its major branches (reviewed in Chapter 1).

The overall goal of the current study is therefore to clarify these higher-order relationships, using an "exemplar taxon" approach (typically using one representative per family), but with a molecular data set derived for each taxon that is greatly expanded compared to previous studies. In the most comprehensive recent studies of the monocots, Chase *et al.* (2000a) noted that some internal branches in Asparagales were short, and suggested that more nucleotide data might help improve the resolution and support for several relationships, including the relationships of Orchidaceae and other basal Asparagales. Fay *et al.* (2000) also suggested that the addition of more data might aid in elucidating the inference of Asparagales phylogeny. I build on their work here by generating a greatly expanded plastid ("chloroplast") data set for a representative collection of Asparagales and relatives.

The regions I survey encompass eight noncontiguous regions in the plastid genome, that span some 20 genes [ten photosystem II genes, three ribosomal proteins genes, three tRNA genes, and genes for two NADH dehydrogenase genes subunits (B and F), the large subunit of ribulose-1, 5-bisphosphate carboxylase/ oxygenase, and subunit B of ATP synthase], and which include a variety of noncoding regions [the *ndh*B, *rps*7, 3'-*rps*12 and *trn*L(UAA) introns, and intergenic spacer regions between the members of the *psb*E-*psb*F-*psb*L-*psb*J operon, between the members of the *psb*B-*psb*T-*psb*N-*psb*H cluster, within the 3'-*rps*12-*rps*7 operon and between this operon and *ndh*B, between *ndh*B and *trn*L(CAA), between *rbc*L and *atp*B, and between *trn*F(GAA) and *trn*L(UAA); see Taberlet *et al.*, 1991, and Graham and Olmstead, 2000a for further details]. I analyze these new data using a variety of methods of phylogenetic analysis to elucidate and clarify higher-order relationships in Asparagales, and the position of the order within the monocots as a whole. I also perform Monte Carlo simulation studies to examine the impact of the putative long-branch problem concerning the placement in Asparagales of the Mediterranean endemic *Aphyllanthes monspeliensis*, the sole species in Aphyllanthaceae.

Materials and Methods

Taxon sampling. Taxa were chosen to exemplify the broad diversity of Asparagales (at the familial level) and outgroups in the monocots (at the ordinal level), based primarily on the classification scheme of APG (1998) and the work of Chase *et al.* (2000a) and Fay *et al.* (2000). Data were obtained from 27 exemplar taxa (Table 3.1) representing 24 of the 29 families currently recognized in Asparagales (APG, 1998; Chase *et al.*, 2000a; Table 1.1). I also included nine exemplars from seven families (representing four orders and an unplaced family) in the commelinoid monocots *sensu* APG (1998), two each from Liliales, Pandanales, and Dioscoreales, three from Alismatales and one from Acorales (*Acorus calamus*). *Acorus* was used to root phylogenetic trees produced here, based on its basal position in broader phylogenetic studies of the angiosperms (e.g., Duvall *et al.*, 1993; Chase *et al.*, 2000a; Graham and Olmstead, 2000a,b; Salvolainen *et al.*, 2000). Source information is provided in Table 3.1. Genera represented in this study are also listed in Table 1.1 (Chapter 1) according to the classification systems of Dahlgren *et al.* (1985), Cronquist (1988), and APG (1998).

DNA extraction, amplification and sequencing. Extraction of genomic DNA was conducted using the basic methodology of Doyle and Doyle (1987; see Graham and Olmstead, 2000a, for further details). Several DNAs were provided by the Royal Botanic Gardens, Kew (see Fay *et al.*, 2000, for methods of DNA extraction). DNA amplification and sequencing were performed using protocols described in Graham and Olmstead (2000a), with the exception that DNA sequencing products were generated using the

"DYEnamic ET terminator cycle sequencing kit" of Amersham Biosciences, following the manufacturer's instructions.

Most of the amplification and sequencing primers used here were designed by Graham and Olmstead (2000a), but additional published primers were employed for the following regions: the *atp*B locus (Hoot *et al.*, 1995); the *ndh*F locus (Olmstead and Sweere, 1994; Kim and Jansen, 1995; Olmstead and Reeves, 1995; Neyland and Urbatsch, 1996 a, b; Graham *et al.*, 1998); the *rbc*L locus (Zurawski *et al.*, 1984); and the intergenic spacer region between *rbc*L and *atp*B (Chiang *et al.*, 1998). I also designed several primers specifically for this study, with the assistance of HS Rai and SW Graham (Table 3.2); four of these are for the intergenic spacer region between *rbc*L and *atp*B, one is for the intergenic spacer region between *ndh*B and *trn*L(CAA), one is for *ndh*F (used for *Amerorchis*), two are for the 3'-*rps*12 and *rps*7 operon, one is for *psb*D, and one is for *rbc*L.

GenBank accession numbers are provided in Table 3.1 for all sequences included in the aligned matrix. I generated most of these for the current study, with some exceptions noted in Table 3.1. These exceptions include several taxa taken in their entirety from completely sequenced plastid genomes available on GenBank (specifically, the three grasses, *Oryza*, *Triticum*, and *Zea*; Table 3.1, footnote 3). Previously published data for the *trn*L-F region [comprising part of *trn*L(UAA), the *trn*L-F intergenic spacer region and a small part of *trn*F(GAA)] were also included for *Lilium superbum* and a subset of Asparagales (Table 3.1, footnote 2). These data were generated by Fay *et al.* (2000; alignment provided by MW Chase and MF Fay, Royal Botanic Gardens, Kew).

Matrix compilation and alignment. Completed DNA sequences ("contigs") were constructed using Sequencher 4.1 (Gene Codes Corp., Ann Arbor, Michigan). These new sequences were added to a previously published alignment (Graham *et al.*, 2002). The final alignment is in NEXUS-format (Swofford, 2002). Details of exon, intron and spacer lengths and boundaries for most of the regions are provided in Graham and Olmstead (2000a). Unaligned lengths of two additional regions included in this study range from 597 bp to 1104 bp (for *Alania endlicheri* Kunth and *Ixiolirion tataricum* (Pall.) Herb. & Traub, respectively; 776 bp in *Xanthorrhoea resinosa* Pers., a reference taxon used below) for the *atp*B-*rbc*L intergenic spacer region, and for the *ndh*B-*trn*L(CAA) intergenic spacer region from 534 bp to 603 bp (for *Ixiolirion tataricum* and *Triticum aestivum* L., respectively, 555 bp in *Xanthorrhoea resinosa*). Individual alignments for each major region were adjusted manually using Se-Al version 1.0 (Rambaut, 1998), following criteria laid out in Graham *et al.* (2000c). The overall alignment was reviewed by more than one person.

Several genes had interrupted reading frames (pseudogenes) for one or a few taxa, or at least had premature stop codons relative to other angiosperm taxa (e.g., compared to those examined in Graham and Olmstead, 2000a, and the rest of this thesis). These sequences were included in all phylogenetic analyses, as exploratory analyses that focussed only on these regions placed the affected taxa in the same position seen with these regions included or excluded (results not shown). Pseudogenes for *ndh*F were retrieved for two of three exemplar taxa from Orchidaceae (*Amerorchis rotundifolium* (Banks) Hultén and *Cypripedium passerinum* Richardson). Two distinct copies of *ndh*B (which included some "intron" and 3'- and 5'-flanking regions) were retrieved from *Cypripedium passerinum* (Table 3.1) using alternative sets of amplification primers. The *ndh*B exons in both cases represent multiply interrupted reading frames, and clearly represent pseudogenes. The first copy of this region includes 3'-*rps*12 and *rps*7 (both open reading frames), the intergenic spacer region between *rps*7, both "exons" of *ndh*B, and a region homologous with the intron of *ndh*B. This copy was used in the final alignment and phylogenetic analysis. The second copy spans a portion of the *ndh*B intron, the second exon of *ndh*B, the intergenic spacer region between *ndh*B and *trn*L(CAA) and a few bases of *trn*L(CAA). The two copies form a clade in parsimony analysis of the 3'-*rps*12-*rps*7-*ndh*B-*trn*L(CAA) region (results not shown), and so it is probable that the duplication that resulted in these copies postdates the separation of *Cypripedium passerinum* from the other orchids sampled here.

Several instances of *ndh* locus loss have been documented in other seed plants. For example, Wakasugi et al. (1994) reported the loss of 11 *ndh* subunit loci in *Pinus*, and Neyland and Urbatsch (1996a) reported probable *ndh*F pseudogenes in at least two clades of orchids. The orchid *ndh*F pseudogenes reported here are not part of either of these clades, and it should be noted that Neyland and Urbatsch (1996a) reported the presence of an *ndh*F open reading frame in *Cypripedium acaule*. As I recovered *ndh*F and *ndh*B pseudogenes from *C. passerinum*, it would be valuable to survey for other *ndh* subunit pseudogenes in other members of *Cypripedium*.

Relative to *Nicotiana tabacum* (GenBank plastid genome accession: Z00044), the *rps*7 sequence obtained for *Lomandra longifolium* Labill. has an 8 bp indel. This indel is 325 bp downstream of the start codon of the gene, and results in a stop codon 109 bp before that seen in *Nicotiana tabacum* (relative to the latter sequence). The predicted translation product for *rps*7 is 119 amino acid residues long in *Lomandra* (compared to 155 residues in *Nicotiana*). It is not known whether this foreshortened *rps*7 is a pseudogene.

It was also not possible to obtain the IGS between *atp*B and *rbc*L for *Cypripedium passerinum*. Preliminary analyses of partial sequences indicate that multiple copies of this region were amplified, similar to the situation with *ndh*B for this taxon. As these copies were incomplete and their relationship to each other is unclear, they were excluded from further analysis. The genomic location(s) of the various foreshortened genes, pseudogenes and duplicated regions is unknown at present.

Phylogenetic analysis of Asparagales.

Plastid data examined. Some taxa had regions with more than 50 bp of contiguous sequence missing (noted in Table 3.1, footnote 1). These regions were coded as "missing data" in the phylogenetic analysis. All data were analyzed except for several short ambiguously aligned regions that were excluded using NEXUS-format character sets (CHARSETs). Excluded regions include short portions of the spacer regions between various photosystem II subunit loci (those between genes in the *psb*E–*psb*F–*psb*L–*psb*J operon and the *psb*B–*psb*T–*psb*N–*psb*H cluster) and the *atp*B-*rbc*L spacer region. Short ambiguously aligned portions of the *trn*L-F region were also omitted during analysis, following exclusion sets defined by Fay *et al.* (2000). The total number of characters considered in the phylogenetic analyses was 21,379 bp aligned nucleotides per taxon, corresponding to 16,728 bp unaligned nucleotides in a reference taxon (*Xanthorrhoea resinosa*). For the same reference taxon, 4,561 bp (unaligned) are from noncoding regions, of which 2,153 bp (unaligned) are from four introns [one each in *ndh*B, *rpl*2, 3'-*rps*12 and *trn*L(UAA)].

Tree search characteristics. All tree searches were conducted using PAUP* version 4.0b10 (Swofford, 2002). For maximum parsimony analysis, all characters and character-state changes were equally weighted, and heuristic searches were conducted using tree-bisection-reconnection (TBR) branch swapping, with 100 random addition replicates. Parsimony-based bootstrap analysis (Felsenstein, 1985) was performed using the same search conditions, except that one random addition replicate was used for each of the 1,000 bootstrap replicates. No tree limit was set for the parsimony analyses. Hillis and Bull (1993) showed that branches with around 70% support tend to be representative of the true phylogeny (over a certain threshold value, bootstrap support values are biased but conservative estimators of support), so long as rates of change are not very high or uneven between lineages. I therefore refer to branches with at least 70% bootstrap support as being relatively robustly supported (see also Graham *et al.*, 1998).

Maximum likelihood (ML) searches were performed for taxon sets that were reduced to facilitate analysis under current computer memory and time limitations. I compared several different likelihood substitution models, and used hierarchical likelihood ratio tests (LRTs; see Huelsenbeck and Crandall, 1997) to pick one of these for subsequent tree-searching. I followed the general LRT protocol described in Huelsenbeck and Crandall (1997), except that I included an additional substitution model (general time-reversible, GTR; Lanave *et al.*, 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez *et al.*, 1990) in the hierarchy of tests, and I also did not test for the presence of a molecular clock. I performed two sets of LRTs. One of these employed as a test tree the single best parsimony tree found using all monocot taxa considered here (Fig. 3.1; see Results). The second employed a parsimony tree that was derived using a taxon set that consisted of all Asparagales surveyed here, plus three outgroups from the commelinoid monocots [*Ananas comosus* (L.) Merr., *Dasypogon hookeri* J. R. Drumm.and *Roystonea princeps* (Becc.) Burret]. This taxon set was employed in subsequent ML analyses. The parsimony tree used for the second set of LRTs was the best tree found in parsimony analyses of these taxa, which is topologically equivalent to the tree employed for the first set of LRTs when that tree is pruned back to these taxa (not shown).

The likelihood model chosen was the same for both model trees: GTR + Γ + I [corresponding to the GTR rate substitution model, with the proportion of invariable sites (I) considered, and with among-site rate variation accounted for using the gamma (Γ) distribution, as described by the rate parameter alpha (α)]. Likelihood model parameters determined for this reduced taxon set were employed in an heuristic search for these taxa, using TBR branch-swapping and "simple" stepwise addition. The same search characteristics and general substitution model were used in subsequent ML searches, but I re-determined model parameters for ML searches when the topology of the model tree for simulation studies (described below) differed from the best ML tree (see Results).

Phylogenetic congruence of different data partitions. The incongruence length difference (ILD) test of Farris *et al.* (1994) was used to gauge the significance of any incongruence among different subset of the data. The test is implemented in PAUP*. It randomly re-partitions the characters from multiple data partitions into new ones as large as the originals (sampling is without replacement). Shortest parsimony trees are determined for each re-partitioned data set, and the sum of tree lengths for the original partitions is compared to the distribution of this sum across all shufflings. If the sum of lengths for the shuffled data sets exceeds that of the unshuffled data for a critical number of re-partitionings, the null hypothesis of congruence is rejected. I repeated the parsimony version of the test (using heuristic searches with TBR branch-swapping and "simple" stepwise addition, for 99 partition replicates) for two data partition comparisons: (i) Codon positions 1 plus 2, versus codon position 3, and; (ii) All coding sequences combined, versus all noncoding regions combined. Data partitions were defined in PAUP* using appropriate CHARSETs; exon and codon position data partitions were determined following the *Nicotiana tabacum* plastid genome (see Graham and Olmstead, 2000a), except that the *Ginkgo* L. start codon was used for *ndh*B. I repeated these tests for all of the monocots considered in this study, and for the members of Asparagales alone.

Further parsimony-based bootstrap analyses were also performed for three subsets of the data (codon positions 1 + 2; codon position 3; all noncoding regions combined). I compared graphically the spectrum of bootstrap support across different taxon partitions (branches) for these data partitions to bootstrap support values determined with all of the data combined (see Graham *et al.*, 1998).

Error and bias in the phylogenetic placement of Aphyllanthaceae in Asparagales. Poor bootstrap support values were obtained for several branches in a clade of "higher Asparagales" (see Chase *et al.*, 1995), particularly those involving Agavaceae, Anthericaceae, Aphyllanthaceae, Hyacinthaceae and Themidaceae. The possibility that this was due to a long-branch problem (Felsenstein, 1978, Hendy and Penny, 1985) associated with Aphyllanthaceae (*Aphyllanthes monspeliensis*) has been suggested by Fay *et al.* (2000). I examined this hypothesis by repeating the main bootstrap analysis with *Aphyllanthes* excluded.

I also used using a Monte Carlo simulation approach, described by Sanderson *et al.* (2000), to examine whether the ambiguous placement of *Aphyllanthes* may be a function of sampling error, or instead may reflect systematic bias. In brief, this approach is performed by taking some tree topology as an "accepted" hypothesis (for instance, the shortest tree), and evolving simulated data sets on this topology, using branch lengths and substitution parameters determined from the real data under some reasonable likelihood substitution model. Phylogenetic analysis of the simulated data sets can then provide conditional estimates of how often the accepted tree topology (or some aspect of it) is inferred, a measure of 1-Type I error (the probability of rejecting the null scenario, if true). The frequency with which a particular topological scenario is inferred when it conflicts with the tree used for the simulations (the "accepted" tree used for simulations then representing an alternative hypothesis), provides a measure of Type II error for the inferred scenario, an estimate that is conditional on the simulation model tree under consideration. Bias is indicated when a skewed subset of one or more trees is recovered, none of which are the model tree.

I used this basic approach to address error and bias, following methodological details noted in Sanderson *et al.* (2000). I performed the Monte Carlo simulations for seven basic topologies (Figure 3.2). Four of these (Figure 3.2, a-d) correspond to suboptimal arrangements of *Aphyllanthes* observed in the bootstrap profile of my main parsimony analysis (whose support ranged from 19-32%). Another is the optimal arrangement in the main analysis (tree "g"; Figure 3.1), and two others (Figure 3.2, e,f) are arrangements observed by Fay *et al.* (2000) using equal weights or successive weights of their data, respectively. For each of these arrangements of *Aphyllanthes*, constrained MP searches were performed with PAUP* (using the "topological constraints" noted in Figure 3.2). Branch lengths and substitution model parameters were determined from each parsimony-based topology, from the real data, using maximum likelihood (for the GTR + Γ + I model). The resulting details and topology (Figure 3.2) were used to generate simulated data sets the same size as the real data, with Seq-Gen, version 1.2.5 (Rambaut and Grassly 1997). MP tree searches for all of the simulated data sets were then performed in batch mode in PAUP*, and the resulting MP trees for each simulated data set were categorized (with scoring checked by a second individual) to determine which of the seven scenarios for *Aphyllanthes* placement (or others, if found) were obtained from each null topology. Following Sanderson *et al.* (2000), other details of tree topology were ignored during tree scoring. When multiple scenarios were inferred for a particular simulation replicate, these were counted as fractional contributions to the total score for the corresponding cell.

Results

Phylogenetic relationships. In the single most-parsimonious tree (Fig. 3.1), the order Asparagales is resolved as the sister group of the commelinoid monocots, with a high bootstrap support value, BV, (93%; upper values on branches). The commelinoids are well supported as a whole (99% BV), as are most relationships within this large group. The orders Alismatales, Dioscoreales, Liliales and Pandanales (all *sensu* APG) are all well supported, at least at their current very coarse taxon samplings (all with 100% BV; Fig. 3.1). However, beyond confirming that Alismatales are a basal monocot order, the interrelationships of the other orders, and their relationship to (Asparagales + commelinoids), are all poorly supported. The shortest parsimony tree, for example, depicts Liliales as the sister group of (Asparagales + commelinoids), with only 53% BV (Fig. 3.1).

Relationships in Asparagales are almost identical in the maximum parsimony analysis and the single tree inferred from maximum likelihood analysis (cf. Figs. 3.1, 3.3). The only difference between them concerns relationships within a clade of three families (Asphodelaceae, Hemerocallidaceae and Xanthorrhoeaceae). While this clade as a whole is well supported by the parsimony-based bootstrap (100% BV), the interrelationships of the three families are not (< 70% BV). In the MP analysis, Hemerocallidaceae and Asphodelaceae are sister taxa (one of two arrangements seen in Chapter 2), whereas in the ML analysis Hemerocallidaceae is the sister group of Xanthorrhoeaceae. In line with Fay *et al.* (2000), the clade consisting of Asphodelaceae-Hemerocallidaceae-Xanthorrhoeaceae is the sister group of the "higher asparagoid" clade of Chase *et al.* (1995), based on the representative families of the latter group sampled here (Fig. 3.1). The family Xeronemataceae is the sister group of this large clade (Fig. 3.1). These relationships are all very robustly supported (at 100% BV). Iridaceae are the sister group of (Xeronemataceae, plus Asphodelaceae-Hemerocallidaceae-Xanthorrhoeaceae, plus the higher Asparagales), a finding noted in some recent studies (e.g., Chase *et al.*, 2000a, Fay *et al.*, 2000), but with strong support here (90% BV). Ixioliriaceae and Tecophilaeaceae form a well-supported clade (91% BV), that is also strongly supported as the sister-group of all of the preceding taxa (100% BV).

Two basal clades of Asparagales have robust support; the orchids (a clade with 100% BV), and the "astelioids," (Chase *et al.*, 1996; Rudall *et al.*, 1998; circumscribed by Fay *et al.*, 2000, to include Boryaceae), a clade with 91% BV. The latter clade (Fig. 3.1) includes all five families hypothesized to belong there by Fay *et al.* (2000). The interrelationships among the astelioid families are also all well supported here (84-100%), in contrast with Fay *et al.* (2000), who found the same clades but with poor support. My analysis corroborates the inclusion of the orchids in Asparagales, with robust support (90% BV; Fig. 3.1), *contra* Dahlgren *et al.* (1985), but in line with APG (1998) and other recent workers. My data also robustly resolve the branching order of the basal-most splits in Asparagales. The orchids are supported as the sister-group of all other Asparagales, and the astelioids (including Boryaceae) are the sister-group of all Asparagales excluding the orchids; all relevant branches supporting these arrangements have at least 90% BV.

Relationships among the families of higher Asparagales are identical to those inferred in Chase *et al.* (2000a), at least concerning families in common to our studies (e.g., excluding Aphyllanthaceae here). With the exception of Agavaceae-Anthericaceae, all these relationships were supported by < 70% BV in Chase *et al.* (2000a; most with < 50% BV), whereas in my study they are all supported by > 75% BV, so long as Aphyllanthaceae is excluded from consideration (Fig. 3.1). My analysis depicts Hyacinthaceae as the sister group of Themidaceae (100% BV), and Anthericaceae as the sister group of Agavaceae (100% BV). These two clades are also sister taxa to each other (with 100% BV), and the four families taken together are the sister group (with good support; 96% BV) of another well supported clade (100% BV) consisting of Convallariaceae, Asparagaceae, and Laxmanniaceae. The latter two families are also resolved as sister groups (with 77% BV). Finally, Alliaceae and Amaryllidaceae are well supported as sister taxa (100% BV), a clade that is also the sister group of the remaining higher Asparagales. My study thus provides further substantial support for the exclusion of Themidaceae from Alliaceae, as proposed by Fay and Chase (1996), Pires *et al.* (2001) and Pires and Sytsma (2002).

While these findings in the higher Asparagales conflict with some relationships in Fay *et al.* (2000), all of the conflicts between our studies involve branches with poor support in their analysis (compare, for example, Figs. 2.3 and 3.1: the data set used in Chapter 2 here is essentially that of Fay *et al.*, 2000). Fay *et al.* (2000) conjectured that the uncertain relationships that they observed within the higher Asparagales were a consequence of the inclusion of *Aphyllanthes monspeliensis* L. in their analysis. This hypothesis is also supported for my data in a Monte Carlo simulation study (see below), but also by large reductions in bootstrap support (22-37%) observed for several branches in the higher Asparagales when *Aphyllanthes* is included in the analysis (compare the upper versus lower branch support values in this part of the tree; Fig. 3.1).

Phylogenetic congruence. The ILD test indicated no significantly heterogeneity across data partitions, of those assessed here. For the first two codon partitions versus the third, $P = 0.06$ (for all taxa) and $P = 0.68$ (for Asparagales alone). For combined coding versus combined noncoding regions, $P = 0.62$ (for all taxa) and $P = 0.66$ (for Asparagales alone). The nearly significant result for the codon position comparison across all taxa should be viewed cautiously, as the test is known to have an excessively high Type I error rate (i.e., it is prone to spurious rejection of the null hypothesis of there being no conflict, particularly if data sets differ in their level of "noise," e.g., Graham *et al.*, 1998; Barker and Lutzoni, 2002).

A comparison of bootstrap support values for branches in and immediately subtending Asparagales (Fig. 3.4) from various data partitions (Fig. 3.5) illustrates that there is very little conflict among these subsets of data. Only branch support values > 50% in at least one of the data partitions were included in Fig. 3.5, for the following four data partitions: (i) all data combined; (ii) the first two codon positions combined; (iii) codon position 3, and; (iv) all noncoding data combined. In general, branches had strongest support when all data were combined, but all the branches of Asparagales depicted in Fig. 3.1 had some bootstrap support from the data subsets considered here. The mean bootstrap support (Olmstead and Sweere, 1994) for branches "a" through "y" in Asparagales (see Fig. 3.4) was highest for codon position 3 (84.8%), compared to a value of 70.9% for the combined noncoding data, and 65.8% for the first two codon positions combined, respectively. (The mean support value for all data combined was 95.0%.) Only a few branches ("ac" through "ae", legend to Fig. 3.5) from individual data partitions are suggestive of substantial incongruence. A clade consisting of Asphodelaceae and Xanthorrhoeaceae ("ac"; Fig. 3.5) has 76% support from codon position 3, but < 50% from all other partitions considered. A clade consisting of the orchids and Liliales ("ad"; Fig. 3.5) has some support from the first two codon positions combined (59%) but < 5% support from all the other partitions. Finally, a clade consisting of Ixioliriaceae and Iridaceae ("ae"; Fig. 3.5) had low support from the noncoding data (54%), but < 50% support from all of the other partitions considered.

Different data partitions also support a variety of positions of *Aphyllanthes* when this is included in bootstrap analysis, but this is unlikely to be a major source of incongruence. The most highly supported position in each case is supported by less than 50% of bootstrap replicates, for all partitions examined (not shown). For example, codon positions 1 and 2 combined support *Aphyllanthes*-Hyacinthaceae (45% BV), codon position 3 supports *Aphyllanthes*-Themidaceae (35% BV), and the noncoding data support *Aphyllanthes*-Alliaceae (44% BV).

Monte Carlo analysis of error and bias in the inference of Aphyllanthes *placement in Asparagales.* Trees inferred from data sets simulated under various scenarios of *Aphyllanthes* placement show mixed patterns of error and bias. If scenario "g" (the observed placement with the real data) or scenario "d" (one of the suboptimal placements seen in the bootstrap analysis of the real data, with 23% BV; Fig. 3.2) were correct, they would both result in correct tree inference, with no error (diagonal elements are 100% in both cases; Table 3.4). Since the actual scenario inferred for simulation "d" is this simulated scenario, and as "g" is the observed scenario, this would appear to reassure us that the observed placement of *Aphyllanthes* with real data (Fig. 3.1) is the correct one. However, if scenario "a" were correct (another suboptimal placements seen in bootstrap analysis, with 32% BV; Fig. 3.2), then there is also a very reasonable chance that the observed scenario, "g," would actually be inferred, incorrectly (45%; Table 3.4). This should be interpreted as systematic bias, conditional on "a" being true, and it means that we cannot determine which, if either, of the two scenarios "a" or "g" is correct, using parsimony analysis alone.

Substantial error estimates are associated with the other major scenarios inferred using the simulated data, including "b," "c," "e" and "f" (the former two were observed in the bootstrap profile derived from the real data, the latter two were observed by Fay *et al.*, 2000; Fig. 3.2). The Type I error for these hypotheses ranges from 11.5 – 37.7% (Table 3.4; diagonal cells). Scenarios "b," "c" and "f" are also occasionally inferred when they are not the tree used for simulations, and the largest conditional Type II error for these scenarios was for scenario "c," which reached nearly 10% when tree "f" was used as the model tree (Table 3.4). Substantial bias is also indicated for most of the model trees considered here, as one or a few scenarios comprised the majority of incorrect inferences in each case (Table 3.4; only simulations "d" and "g" had no indication of any bias). Although none of scenarios (apart from "a" and "g") resulted in the inference of the scenario observed with the real plastid data (i.e., with *Aphyllanthes* depicted as the sister group of Agavaceae-Anthericaceae; Fig. 3.1), we obviously cannot rule out the possibility that other scenarios that were not tested could also have yielded the observed placement of *Aphyllanthes*.

Discussion

With some noteworthy exceptions, the higher-order relationships depicted in my parsimony and likelihood based trees are highly congruent with those observed in the most comprehensive recent studies of Asparagales, those of Chase *et al.* (1995) and Fay *et al.* (2000), and also with two studies with fewer representatives of Asparagales (Chase *et al.*, 2000a; Rudall *et al.*, 1997). However, the relationships generally have substantially stronger bootstrap support from the large plastid data set examined here than in those studies (cf. Fig. 2.3 and Fig. 3.1, for example).

My study provides definitive support for: (i) the sister-group relationship between Asparagales and the commelinoid monocots; (ii) the inclusion of Orchidaceae in Asparagales, and; (iii) the branching order at the base of Asparagales. All of these arrangements come with at least 90% bootstrap support (Fig. 3.1). These findings have been noted elsewhere, but generally with poor support. The only other study that provided comparable support for the first of these two findings, that of Fay *et al.* (2000), included outgroup representatives from only four orders; Liliales, Pandanales and two orders of commelinoid monocots.

I robustly resolved additional major sets of relationships that are poorly supported or contradictory in other studies, including the relationships among a group of five small families, the "astelioids" (Fig. 3.1), found with poor support by Fay *et al.* (2000), and hinted at, in part, by Chase *et al.* (1995). I found the same relationship among these five families as Fay *et al.* (2000). The family Asteliaceae is the sister group of Lanariaceae-Hypoxidaceae, Blandfordiaceae are the sister group of these three, and Boryaceae are part of the basal split in this astelioid clade. All of the relationships among these families find robust support here (84-100% BV).

A sister-group relationship between Ixioliriaceae and Tecophilaeaceae was well supported here and has been observed in some recent molecular studies (Chase *et al.*, 2000a; Fay *et al.*, 2000, but not in Chase *et al.*, 1995). Doryanthaceae may be the actual sister-group of Ixioliriaceae, or it may be located nearby along the major backbone of Asparagales relationships (variable relationships have been observed within the same study; Fay *et al.*, 2000, and see Chapter 2). Doryanthaceae will obviously be an important taxon to add in future expansions of the data examined here.

My findings uphold the relationships among the families of the higher Asparagales found by Chase *et al.* (2000a), in contrast to those observed by Fay *et al.* (2000). Of particular note, I find Hyacinthaceae and Themidaceae to be sister taxa (found by Chase *et al.*, 2000a, and Pires *et al.*, 2002), and also find Asparagaceae and Laxmanniaceae to be each others closest relatives. Both findings are robustly supported. In their successively weighted analysis, Fay *et al.* (2000) found Convallariaceae to be the sister-group of Asparagaceae, and Themidaceae to be the sister-group of a large cluster of families including Agavaceae, Anthericaceae, Aphyllanthaceae and Hyacinthaceae. These findings conflict with my results and those of Chase *et al.* (2000a), but do not involve branches that are well supported in Fay *et al.* (2000); their results are likely spurious. This may be at least in part a function of the inclusion of *Aphyllanthes monspeliensis* in their phylogenetic analysis, and indeed Fay *et al.* (2000) hypothesized that *Aphyllanthes* destabilized their phylogenetic inference in the higher Asparagales, as they found different local arrangements in successively weighted versus flat-weighted analyses. This seems to hold true here too, although to a more limited extent: three branches show substantial drops in bootstrap support with *Aphyllanthes* included (Fig. 3.1).

I find good bootstrap support (>70%) throughout the rest of the higher Asparagales, even with *Aphyllanthes* included, although it is not well resolved where exactly *Aphyllanthes* fits among Agavaceae, Anthericaceae, Hyacinthaceae and Themidaceae (Fig. 3.1). Whether a 75% bootstrap value provides robust support for the local inclusion of *Aphyllanthes* in a clade with Agavaceae, Anthericaceae, Hyacinthaceae and Themidaceae is unclear, as bootstrap analysis may be misleading when there are problematic "long branches" (*sensu* Felsenstein, 1978, Hendy and Penny, 1985); see Hillis and Bull (1993). Indeed, the Monte Carlo simulation analysis suggests that parsimony analysis may be strongly misleading concerning the placement of *Aphyllanthes*, if its true position is as the sister group of Hyacinthaceae-Themidaceae (Fig. 3.2; Table 3, simulation "a"), or perhaps elsewhere (see Results). However, maximum likelihood analysis is expected to be less prone to long branch effects, particularly if the substitution model used for analysis bears a strong resemblance to the real evolutionary model (Chang, 1996; Swofford et al., 2001). The ML placement of *Aphyllanthes* depicts it in the same position as the MP analysis, as the sister group of Anthericaceae-Agavaceae (Figs. 3.1, 3.3). It would be valuable to repeat the Monte Carlo simulation study using ML inference of the simulated data, instead of MP, to determine how prone the former analytical method is to being mislead in this particular instance.

There may be other undetected long branches in Asparagales phylogenetic inference, but this seems doubtful. *Aphyllanthes* was suggested to be a problem taxon by

Fay *et al.* (2000) based on its tendency to destabilize branches locally. Most of the other branches in Asparagales have extremely robust support here (Fig. 3.1). One issue that may warrant further investigation concerns the relationships among Asphodelaceae, Hemerocallidaceae and Xanthorrhoeaceae. The relatively poor bootstrap support observed for the MP relationships among these three taxa (Fig. 3.1) may represent another long-branch problem, or it may simply be a consequence of sampling error alone (too few characters). As the branch supporting the basal split in Asphodelaceae-Hemerocallidaceae-Xanthorrhoeaceae is extremely short (Fig. 3.3), the latter explanation seems more likely.

In general, there is not expected to be strong systematic bias in the inference of phylogenetic relationships from plastid data, even at relatively deep levels of phylogeny. Savolainen *et al.* (2002) demonstrated that the very strong systematic biases observed during inference of deep animal phylogenies from mitochondrial data (Naylor and Brown, 1998) are unlikely to hold, or to have much effect on phylogenetic inference, for the two most commonly used plastid genes in plant systematics work (*atp*B and *rbc*L; both used here). I expect that this is equally true for the regions that I added to the analysis of Asparagales phylogenetics here, for two reasons. First, these regions are all very slowly evolving, including most of the noncoding regions employed in the phylogenetic analysis here (see Graham and Olmstead, 2000a, b; Graham *et al.*, 2002);

slowly evolving characters are expected to be less prone to long-branch problems than more rapidly evolving characters (Felsenstein, 1983). Second, the ILD test of Farris *et al.* (1994) indicated no evidence of incongruence, at least within Asparagales, for several data partitions that evolve at quite different rates (different codon position comparisons, coding versus noncoding data). Bootstrap analyses of these different data partitions yield highly congruent results (Fig. 3.5), and although support for particular branches is often substantially lower than that observed with all of the data combined, there are practically no robustly supported branches from analysis of individual partitions that contradict those seen in the other data partitions, or the main analysis with all data combined. One possible exception is branch "ac" (representing Asphodelaceae-Xanthorrhoeaceae), supported by codon position 3 data, but not substantially by any other data set (Fig. 3.5). However, the latter arrangement may simply be a consequence of sampling error (see above).

Morphological evidence of higher-order relationships. The findings reported here for molecular characters support most of the major results of other recent studies on Asparagales, although with better support. However, there is a dearth of nonmolecular evidence for many of these groupings, and indeed of macromorphological characters to support many of the current familial circumscriptions (Fay *et al.*, 2000). A future focus of research in the order will therefore be to uncover morphological and anatomical characters that support many of the higher-order relationships inferred from molecular data. The existence of well supported phylogenies, such as the one reported here, should further promote this search.

This type of search has already yielded results (reviewed in Fay *et al.*, 2000 and Judd *et al.*, 2002), including the recognition of several possible synapomorphies for the order, such as simultaneous microsporogenesis (Rudall et al., 1997; but note that this character has reversed to successive microsporogenesis in the higher Asparagales, Xanthorrhoeaceae and Hypoxidaceae). Fay *et al.* (2000) also suggest that inferior ovaries are a synapomorphy for Asparagales, although ovary position has apparently undergone multiple changes in the order [the clade (higher Asparagales + Xeronemataceae) is supported by a superior ovary, and there have apparently been subsequent reversals to inferior ovaries in Amaryllidaceae and some Agavaceae (Fay *et al.*, 2000)].

The clade consisting of Asphodelaceae, Hemerocallidaceae and Xanthorrhoeaceae is supported by the synapomorphy of anthraquinone possession (Kite *et al.*, 2000). A clade consisting of Alliaceae-Amaryllidaceae and Agapanthaceae is supported by bulbs and the presence of an umbellate inflorescence (Fay *et al.*, 2000; Judd *et al.*, 2002), although Themidaceae have apparently evolved umbels convergently, contributing to their onion-like appearance.

Outstanding questions in Asparagales higher-order systematics. Because I focussed on collecting very large amounts of data per taxon, I had to leave some taxa out of this study that are undoubtedly part of the major backbone of Asparagales relationships. Future research will focus on including these taxa. The most outstanding unresolved questions involve the placement of Doryanthaceae (likely a lineage emerging along the backbone subtending the higher Asparagales) and Agapanthaceae (whose precise relationship to Alliaceae and Amaryllidaceae is currently unresolved; Fay *et al.*, 2000). All of the families of Asparagales *sensu* APG that I did not sample should be included in future work on the order. Also missing here are Anemarrhenaceae, Behniaceae, Herreriaceae and Hesperocallidaceae; these likely all belong in the higher Asparagales (Fay *et al.*, 2002). Including multiple representatives per family would also be valuable, to help break up some of the remaining long branches in the order, and to help clear up some deep and unresolved phylogenetic relationships in the larger families of Asparagales, such as Iridaceae and Orchidaceae.

Table 3.1. GenBank accession numbers and vouchers for exemplar monocot taxa [1, 2, 3]

	Gene or region								
Taxon [Voucher (herbarium)]	*atp*B	*ndh*F [4]	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L [5]	IGS (*atp*B- *rbc*L) [6]
ACORALES									
Acorus *calamus* L. [see Graham and Olmstead, 2000a]	(AJ235381) [7]	AY007647 [7]	AF123843 [7]	AF123813 [7]	AF123828 [7]	(D28865) [7]	AF123785 [7]	AF123771 [7]	n/a
ALISMATALES									
Sagittaria *latifolia* Willd. [SCH Barrett s.n. (TRT)]	AF239788	AY007657 [8]	AY007469	AF239789	AY007484	(L08767) [7]	AY007497	AF238074	n/a
Scheuchzeria *palustris* L. [M Waterway & SW Graham 97-60 (ALTA)]	AY147594	AF547007	AY147500	AY147639	AY147547	(U03728) [7]	AY147686	AY147451	n/a
Spathiphyllum *wallisii* Hort. [MW Chase 210 (NCU)]	(AJ235606.2) [7]	AY007658 [8]	AY007471	AF239794	AY007487	(AJ235807) [7]	AY007500	AF238077	n/a
ASPARAGALES									
Alania *endlicheri* Kunth [DH Vitt 27706 (ALTA)]	AY147612	AY147773	AY147519	AY147658	AY147566	(Y14982) [7]	AY147705	AY147471	AY147737

Taxon [Voucher (herbarium)]	Gene or region								
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L	IGS (*atp*B- *rbc*L)
ASPARAGALES (contd.)									
Allium *textile* A. Nelson & J. F. MacBride [MA McPherson 990704-79 (ALTA)]	AY147628	AF547000	AY147536	AY147675	AY147583	AY149372	AY147721	AY147489	AY147753
Amerorchis *rotundifolia* (Banks) Hultén [MA McPherson 010610-1 (ALTA)]	AY147623	AY147783 [4]	AY147531	AY147670	AY147578	AY149368	AY147716	AY147484	AY147748
Aphyllanthes *monspeliensis* L. [SW Graham 00-04-08 (ALTA)]	AY147629	AY147787	AY147537	AY147676	AY147584	AY149373	AY147722	AY147490	AY147754
Asparagus *officinalis* L. [MA McPherson 010819-2 (ALTA)]	AY147630	AY147788	AY147538	AY147677	AY147585	AY149374	AY147723	AY147491	AY147755
Asphodelus *albus* Willd. [L Harder 1-000430 (ALTA)]	AY147613	AY147774	AY147520	AY147659	AY147567	AY149360	AY147706	AY147472	AY147738
Astelia *alpina* R. Br.ssp. *alpina* [MW Chase 1103 (NCU)]	AY147614	AY147775	AY147521	AY147660	AY147568	(Z77261) [7]	AY147707	AY147473	AY147739
Blandfordia *punicea* (Labill.) Sweet [MW Chase 519 (NCU)]	AY147615	AY147776	AY147522	AY147661	AY147569	(Z73694) [7]	AY147708	AY147474	AY147740
Chlorophytum *comosum* (Thunb.) Jacques [MA McPherson 000321-1 (ALTA)] [9]	AY147631	AY147789	AY147539	AY147678	AY147586	AY149375	AY147724	AY147492	AY147756

Taxon [Voucher (herbarium)]	Gene or region								
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L	IGS (*atp*B- *rbc*L)
ASPARAGALES (contd.)									
Coelogyne cristata Lindl. [MA McPherson 010921-1 (ALTA)] [9]	AY147616	AY147777	AY147523	AY147662	AY147570	AY149361	AY147709	AY147475	AY147741
Curculigo capitata (Lour.) Kuntze [MW Chase 205 (NCU)]	AY147617	AY147778	AY147524	AY147663	AY147571	AY149362	AY147710	AY147476	AY147742
Cyanastrum cordifolium Oliver [Graham & Barrett 2 (TRT)]	(AF168902) [7]	U79228 [8]	AY147525	AY147664	AY147572	U41572 [8]	AY147711	AY147477	AY147743
Cypripedium passerinum Richardson [MA McPherson 010722-6 (ALTA)]	AY147618	AY147779 [4]	AY147526	AY147665	AY147573	AY149363	AY147712	AY147478 [10] AY147479 [10]	n/a
Hemerocallis littorea Makino [MW Chase 3833 (K)]	AY147619	AY147780	AY147527	AY147666	AY147574	AY149364	AY147713	AY147480	AY147744
Iris missouriensis Nutt. [MA McPherson 000707-5a-7 (ALTA)]	AY147620	AF547003	AY147528	AY147667	AY147575	AY149365	AY147714	AY147481	AY147745
Ixiolirion tataricum (Pall.) Herb. & Traub [MW Chase 489 (K)]	AY147621	AY147781	AY147529	AY147668	AY147576	AY149366	n/a	AY147482	AY147746

Taxon [Voucher (herbarium)]	Gene or region								
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L	IGS (*atp*B- *rbc*L)
ASPARAGALES (contd.)									
Lanaria *lanata* (L.) Druce [MW Chase 458 (NCU)]	AY147622	AY147782	AY147530	AY147669	AY147577	AY149367	AY147715	AY147483	AY147747
Lomandra *longifolium* Labill. [DH Vitt 27411 (ALTA)]	AY147632	AF547004	AY147540	AY147679	AY147587	(L05039.2) [7]	AY147725	AY147493	AY147757
Maianthemum *racemosa* (L.) Link [MA McPherson 990704-97 (ALTA)]	AY147633	AF547005	AY147541	AY147680	AY147588	AY149376	AY147726	AY147494	AY147758
Muilla *maritima* S. Watson [JC Pires 98-028 (WIS)]	AY147634	AY147790	AY147542	AY147681	AY147589	AY149377	AY147727	AY147495	AY147759
Muscari *comosum* (L.) Miller [L Harder 000419-1 (ALTA)]	AY147635	AF547006	AY147543	AY147682	AY147590	AY149378	AY147728	AY147496	AY147760
Narcissus *elegans* (Haw.) Spach [SCH Barrett 1434 (TRT)]	AY147636	U79216 [8]	AY147544	AY147683	AY147591	AY149379	AY147729	AY147497	AY147761
Phormium *tenax* J. R. Forst. & G. Forst. [MA McPherson 000612-3 (ALTA)]	AY147624	AY147784	AY147532	AY147671	AY147579	(Z69232) [7]	AY147717	AY147485	AY147749

Taxon [Voucher (herbarium)]	Gene or region								
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L	IGS (*atp*B- *rbc*L)
ASPARAGALES (contd.)									
Sisyrinchium montanum Greene [MA McPherson 990704-71 (ALTA)]	AY147625	AF547008	AY147533	AY147672	AY147580	AY149369	AY147718	AY147486	AY147750
Xanthorrhoea resinosa Pers. [MW Chase 192 (NCU)]	AY147626	AY147785	AY147534	AY147673	AY147581	AY149370	AY147719	AY147487	AY147751
Xeronema callistemon W. R. B. Oliv. [MW Chase 653 (K)]	AY147627	AY147786	AY147535	AY147674	AY147582	AY149371	AY147720	AY147488	AY147752
Yucca glauca Nutt. [Voucherless field collection (Transect #6, Onefour, Alberta; J. Addicott & D. Hurlburt). (*SWG 00121* DNA)]	AY147637	AF547014	AY147545	AY147684	AY147592	AY149380	AY147730	AY147498	AY147762
COMMELINALES									
Ensete ventricosum (Welw.) Cheesman [Kress 96-5372 (US)]	(AF168910)[7]	AY147769	AY147510	AY147649	AY147557	AY149354	AY147696	AY147461	AY147733
Philydrum lanuginosum Banks and Sol. ex Gaertn. [Graham & Barrett 1 (TRT)]	AY147607	U41622[8]	AY147514	AY147653	AY147561	U41596[8]	AY147700	AY147465	AY147734

Taxon [Voucher (herbarium)]	Gene or region								
	*atp*B	*ndh*F	*psb*B, T, N, & *psb*H	*psb*D & C	*psb*E, F, L & *psb*J	*rbc*L	*rpl*2	3'-*rps*12, *rps*7, *ndh*B & *trn*L	IGS (*atp*B- *rbc*L)
COMMELINALES (contd.)									
Roystonea princeps (Becc.) Burret [E Santiago #J-4, (UPR)]	AY147608	AY147772	AY147515	AY147654	AY147562	AY149357	AY147701	AY147466	AY147735
DIOSCOREALES									
Burmannia capitata Mart. [R Neyland 958 (MCN)]	AY147596	n/a	AY147502	AY147641	AY147549	AY149347	AY147688	AY147453	n/a
Dioscorea bulbifera L. [see Graham and Olmstead, 2000a]	AF187059 ^{7}	AY007652 ^{7}	AF123849 ^{7}	AF123819 ^{7}	AF123834 ^{7}	(D28327) ^{7}	AF123791 ^{7}	AF123777 ^{7}	n/a
LILIALES									
Anticlea elegans (Pursh) Rydberg (= *Zigadenus elegans* Pursh) [MA McPherson 990704-68 (ALTA)]	AY147600	AY147765	AY147506	AY147645	AY147553	AY149351	AY147692	AY147457	n/a
Lilium superbum L. [MW Chase 112 (NCU)]	AY116649	AY007655 ^{8}	AY007465	AF239783	AY007480	(L12682) ^{7}	AY007493	AF238070	n/a

Taxon [Voucher (herbarium)]	Gene or region *atp*B	 *ndh*F	 *psb*B, T, N, & *psb*H	 *psb*D & C	 *psb*E, F, L & *psb*J	 *rbc*L	 *rpl*2	 3'-*rps*12, *rps*7, *ndh*B & *trn*L	 IGS (*atp*B- *rbc*L)
POALES [3]									
Ananas comosum (L.) Merr. [HS Rai 1003 (ALTA)]	AY147601	AY147766	AY147507	AY147646	AY147554	(L19977) [7]	AY147693	AY147458	AY147731
Typha latifolia L. [MA McPherson 010819-3 (ALTA)]	AY147610	n/a	n/a	n/a	n/a	(M91634) [y]	AY147703	AY147469 [11]	AY147736
Typha angustifolia L. [Graham 1040 (ALTA)]	n/a	U79230 [8]	AY147517	AY147656	AY147564	n/a	n/a	AY147468 [11]	n/a
PANDANALES									
Stemona tuberosa Lour. [Rothwell & Stockey 46 (ALTA)]	AY147599	AF547009	AY147505	AY147644	AY147552	AY149350	AY147691	AY147456	n/a
Talbotia elegans Balf. [Rothwell & Stockey 48 (ALTA)]	AY147609	AF547011	AY147516	AY147655	AY147563	AY149358	AY147702	AY147467	n/a
UNCERTAIN ORDINAL PLACEMENT									
Dasypogon hookeri J. R. Drumm. [M. W. Chase 430 (NCU) & Conran *et al.* 917 (PERTH)] [12]	AY147603	AY147768	AY147509	AY147648	AY147556	AY149353	AY147695	AY147460	AY147732

[1] Partly sequenced genes and noncoding sequences by taxon, relative to sequences considered in Graham and Olmstead (2000a); minor missing portions (<50bp) at the ends of major regions) are not listed: (1) *psb*D: *Ixiolirion*, 849 bp lacking at 3'-end; (2) *rbc*L: all taxa lack 196 bp at the 3'-end except *Lomandra* and *Roystonea*; (3) *atp*B: *Asparagus* and *Astelia* missing 63 bp and 732 bp at the 3'-end, respecitively; (4) *ndh*F: *Burmannia* not included; *Hemerocallis*, *Orchis*, and *Cypripedium* missing 562 bp, 566 bp, 1249 bp, respectively, at the 3'-end (4) *rpl*2: *Ixiolirion* not included; *Lanaria* missing 89 bp at 3'-end; (5) 3'-*rps*12: *Talbotia* missing 59 bp from 5'-end; the 3'-*rps*12 intron has been lost in *Asphodelus* and *Hemerocallis* (see Chapter 2).

[2] Data from Fay *et al.* (2000) were included in the main alignment from the *trn*L(UAA)-*trn*F(GAA) region for a subset of Asparagales (see Fay *et al.* for source information). These sequences were derived from the same or different source specimens for the particular exemplar species included here (for *Alania*, *Asparagus*, *Aphyllanthes*, *Astelia*, *Lilium*, *Hemerocallis*, *Ixiolirion*, *Muilla*, *Xanthorrhoea* and *Xeronema*). In several cases different species from the same genus (*Allium altaicum* Pall., *Asphodelus aestivus* Brot. , *Cypripedium irapeanum* La Llave & Lex., *Iris unguicularis* Poir.) or different genera from the same family (*Kabuyea hostifolia* (Engl.) Brummitt for *Cyanastrum* of Tecophilaeaceae and *Thysanotus spiniger* Brittan for *Lomandra* of Laxmanniaceae) were used by Fay *et al.* (2000), and here also.

[3] The three grasses included here (*Oryza sativa* L., *Triticum aestivum* L. and *Zea mays* L.) are from whole chloroplast genomes deposited on GenBank (accessions NC_001320, NC_002762 and NC_001666.2, respectively).

[4] Pseudogene in *Amerorchis* and *Cypripedium*.

[5] In a change to a previously published matrix (Graham and Olmstead 2000a) the intergenic spacer region between *ndh*B and *trn*L(CAA) was included for all exemplars examined here from Arecales, Asparagales (except *Cypripedium passerinum*), Commelinales, Poales, Zingiberales, *Dasypogon hookeri* and *Stemona tuberosa*.

[6] In a change to a previously published matrix (Graham and Olmstead 2000a) the intergenic spacer region between *atp*B and *rbc*L was included for all taxa examined here from Arecales, Asparagales (except *Cypripedium passerinum*), Commelinales, Poales, Zingiberales and *Dasypogon hookeri*).

[7] Previously published sequences. Accessions in round brackets were produced by other laboratory groups; some of these are derived from different source specimens for the same species.

[8] Sequence were updated (lengthened) compared to versions used in Graham and Olmstead (2000) and Graham *et al.* (2002).

[9] Cultivars.

[10] The full second exon of *ndh*B was also included for these taxa (ca. 74 additional bp at the 3'-end). Two pseudogenes (and associated regions) were retrieved from *Cypripedium passerinum* for *ndh*B; the sequence represented by the first GenBank accession number (AY147478) was used in the analysis. This excludes this extra portion of *ndh*B, and the intergenic spacer region between *ndh*B and *trn*L(CAA). The second *Cypripedium ndh*B pseudogene fragment (AY147479) comprises a 528 bp fragment homologous to the intron, all of exon 2 and the intergenic spacer region between this *ndh*B pseudogene and *trn*L(CAA).

[11] The matrix sequence is a composite of 3'-*rps*12, *rps*7, the first exon *ndh*B and intron of *ndh*B (from *Typha angustifolia*), and the second exon of *ndh*B and the intergenic spacer between *ndh*B and *trn*L (from *Typha latifolia*).

[12] *Dasypogon hookeri* sequences are composites from two sources of DNA from the same species. No conflicts were observed for several short regions where both sources contributed to the final contigs.

Table 3.2. Primers designed for the current study.

Genomic region	Primer name	[1] Primer sequence
Spacer region between *atp*B and *rbc*L		
	IGS1F	5'- GGATTTACATATACAACATATACCACTGTCA
	IGS2R	5'- GACAGTGGTATATGTTGTATATGTAAATCC
	IGS3R	5'- TTGTATACTCGTTGATATATATGGCG
	IGS4F	5'- GTTTCATTTCTCTATCATTATA
Spacer region between *ndh*B and *trn*L(CAA)		
	B14F	5'- CACCAGAATAGATAAAGTTTTCC
*ndh*F		
	[2] ModNey-7F	5'- GGYACACTTTCTCTTTGYGG
3'-*rps*12 and *rps*7		
	B3R	5'- GATTGGAAATCRTGTATTTTC
	B5F	5'- CGTATTCTTAAACACGGAAAAAAATC
*psb*D		
	C40F	5'- CTATTGCTYTTTCCTTGYGCTTATTTTGC
*rbc*L		
	B83R	5'- AAATCAAGTCCACCRCGTAGACATTC

[1] Degenerate sites follow IUPAC/ IUB ambiguity codes.

[2] Modification of primer 7 of Neyland and Urbatsch (1996).

Table 3.3. Hierarchical likelihood ratio tests (all data combined) comparing different substitution models. The first set of tests (A) used the single most parsimonious (MP) tree topology (Fig. 3.1); the second set (B) used the best MP tree determined for Asparagales and three outgroup taxa (see text). For the degrees of freedom for each test see Table 2.2 (Chapter 2).

Substitution model	-ln likelihood	Comparison[1]	$-2 \ln \Lambda$ [2]	P [3]
(A)				
JC69	129,469.98	-----	-----	-----
F81	128,647.09	JC69 vs. F81	822.89	< 0.01
HKY85	124,209.88	F81 vs. HKY85	4,437.21	< 0.01
GTR	123,763.68	HKY85 vs. GTR	446.20	< 0.01
GTR + Γ	114,085.48	GTR vs. (GTR + Γ)	9,678.2	< 0.01
GTR + Γ + I	113,704.65	(GTR + Γ) vs. (GTR + Γ + I)	380.83	< 0.01
(B)				
JC69	82,785.28	-----	-----	-----
F81	82,031.61	JC69 vs. F81	753.67	< 0.01
HKY85	79,547.84	F81 vs. HKY85	2,483.77	< 0.01
GTR	79,289.35	HKY85 vs. GTR	258.49	< 0.01
GTR + Γ	75,226.09	GTR vs. (GTR + Γ)	4,063.26	< 0.01
GTR + Γ + I	74,997.65	(GTR + Γ) vs. (GTR + Γ + I)	228.44	< 0.01

[1] Abbreviations: JC69 = Jukes Cantor (1969); F81 = Felsenstein (1981); HKY85 = Hasegawa *et al.* (1985); GTR = General Time-Reversible (Lanave *et al.*, 1984; Tavaré, 1986; Barry and Hartigan, 1987; Rodríguez *et al.*, 1990). Γ = Gamma; I = Proportion of invariable sites.

[2] Likelihood ratio test statistic.

[3] The α-level was adjusted using a Bonferroni correction for ten tests.

Table 3.4. Estimated error rates for maximum parsimony inference given specified tree topologies. Six of the seven trees (a-f) used to simulate data are shown in Fig. 3.2; the last (g) is the optimal parsimony tree (Fig. 3.1)

	Scenarios inferred from simulated data:							
	a	b	c	d	e	f	g	Others [1]
Trees used for simulations:								
Tree a	**0.520**	0.015	0.020	0	0	0	0.445	0
Tree b	0	**0.885**	0.010	0	0	0	0	0.105
Tree c	0	0.070	**0.775**	0	0	0.005	0	0.150
Tree d	0	0	0	**1.000**	0	0	0	0
Tree e	0	0	0	0	**0.713**	0	0	0.287
Tree f	0	0	0.098	0	0	**0.623**	0	0.283
Tree g	0	0	0	0	0	0	**1.000**	0

[1] Arrangements of *Aphyllanthes* and relatives that are not present in scenarios "a" through "g". These were predominantly of one tree topology for each simulation. The majority placement of *Aphyllanthes* in the "others" category was as follows (with proportions noted):

Simulation b – ((*Aphyllanthes*, *Muilla*), ((*Chlorophytum*, *Yucca*), *Muscari*)); 0.095
Simulation c – ((*Aphyllanthes*, *Muscari*), ((*Chlorophytum*, *Yucca*), *Muilla*)); 0.145
Simulation e – ((*Aphyllanthes*, *Lomandra*), (*Asparagus*, *Maianthemum*)); 0.213
Simulation f - ((*Aphyllanthes*, *Muscari*), ((*Chlorophytum*, *Yucca*), *Muilla*)); 0.283.

NOTE. -- Trees in the left-most column were used to simulate the data sets (see text and Fig. 3.2). Diagonal cells (bold) are 1-Type I error estimates, and off-diagonal cells are Type II error estimates for each inferred scenario.

FIG. 3.1. Single most-parsimonious tree inferred for Asparagales and related monocots based on a large chloroplast data set [including *atp*B, *ndh*B, *ndh*F, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, *psb*T, *rbc*L, *rpl*2, *rps*7, 3'-*rps*12, *trn*L(UAA) and associated introns and other noncoding regions]. Tree length = 16,889 steps, consistency index = 0.493, retention index = 0.480. Bootstrap values below branches are from an analysis with all taxa included; values above are with *Aphyllanthes monspeliensis* excluded (see text). The commelinoid monocots and two large clades within Asparagales (the astelioids and higher Asparagales) are highlighted.

Figure 3.1

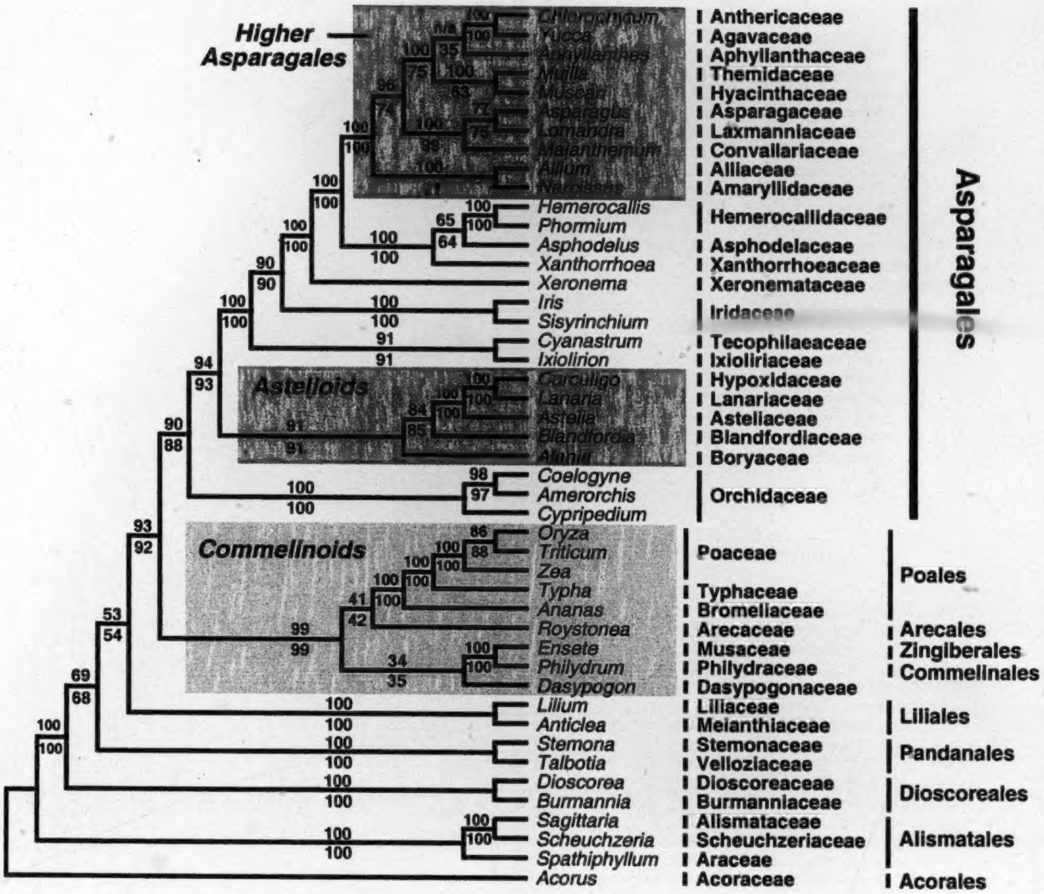

Higher Asparagales
Chlorophytum
Yucca
Aphyllanthes
Muilla
Muscari
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Hemerocallis
Phormium
Asphodelus
Xanthorrhoea
Xeronema
Iris
Sisyrinchium
Cyanastrum
Ixiolirion
Asteliods
Curculigo
Lanaria
Astelia
Blandfordia
Alania
Coelogyne
Amerorchis
Cypripedium
Commelinoids
Oryza
Triticum
Zea
Typha
Ananas
Roystonea
Ensete
Philydrum
Dasypogon
Lilium
Anticlea
Stemona
Talbotia
Dioscorea
Burmannia
Sagittaria
Scheuchzeria
Spathiphyllum
Acorus
Anthericaceae
Agavaceae
Aphyllanthaceae
Themidaceae
Hyacinthaceae
Asparagaceae
Laxmanniaceae
Convallariaceae
Alliaceae
Amaryllidaceae
Hemerocallidaceae
Asphodelaceae
Xanthorrhoeaceae
Xeronemataceae
Iridaceae
Tecophilaeaceae
Ixioliriaceae
Hypoxidaceae
Lanariaceae
Asteliaceae
Blandfordiaceae
Boryaceae
Orchidaceae
Poaceae
Typhaceae
Bromeliaceae
Arecaceae
Musaceae
Philydraceae
Dasypogonaceae
Liliaceae
Melanthiaceae
Stemonaceae
Velloziaceae
Dioscoreaceae
Burmanniaceae
Alismataceae
Scheuchzeriaceae
Araceae
Acoraceae
Asparagales
Poales
Arecales
Zingiberales
Commelinales
Liliales
Pandanales
Dioscoreales
Alismatales
Acorales

FIG. 3.2. Portions of the trees used for Monte Carlo simulations, represented as parsimony phylograms. The trees are the shortest ones from parsimony searches that constrained *Aphyllanthes monspeliensis* in various configurations. The length increase found in these constrained searches compared to the MP tree (Fig. 3.1) is noted beside each tree. Fig. 3.2, a-d; the top four suboptimal positions of *Aphyllanthes monspeliensis* in bootstrap analysis (32%, 19%, 20% and 23%, respectively). Fig. 3.2, e; the local configuration of *Aphyllanthes* and relatives found by Fay *et al*. (2000) using flat-weighted parsimony. Fig. 3.2, f; the placement of *Aphyllanthes* found by Fay *et al*. (2000) using successive weighting. A seventh tree (g, Fig. 3.1; *Aphyllanthes* placement supported by 35% of bootstrap replicates) is the optimal tree found in unconstrained analysis of my data (Fig. 3.1). The shaded regions of the trees show the taxa constrained during the search to obtain the tree. The portions of the trees not shown are the same across all seven trees (see Fig. 3.1).

Figure 3.2

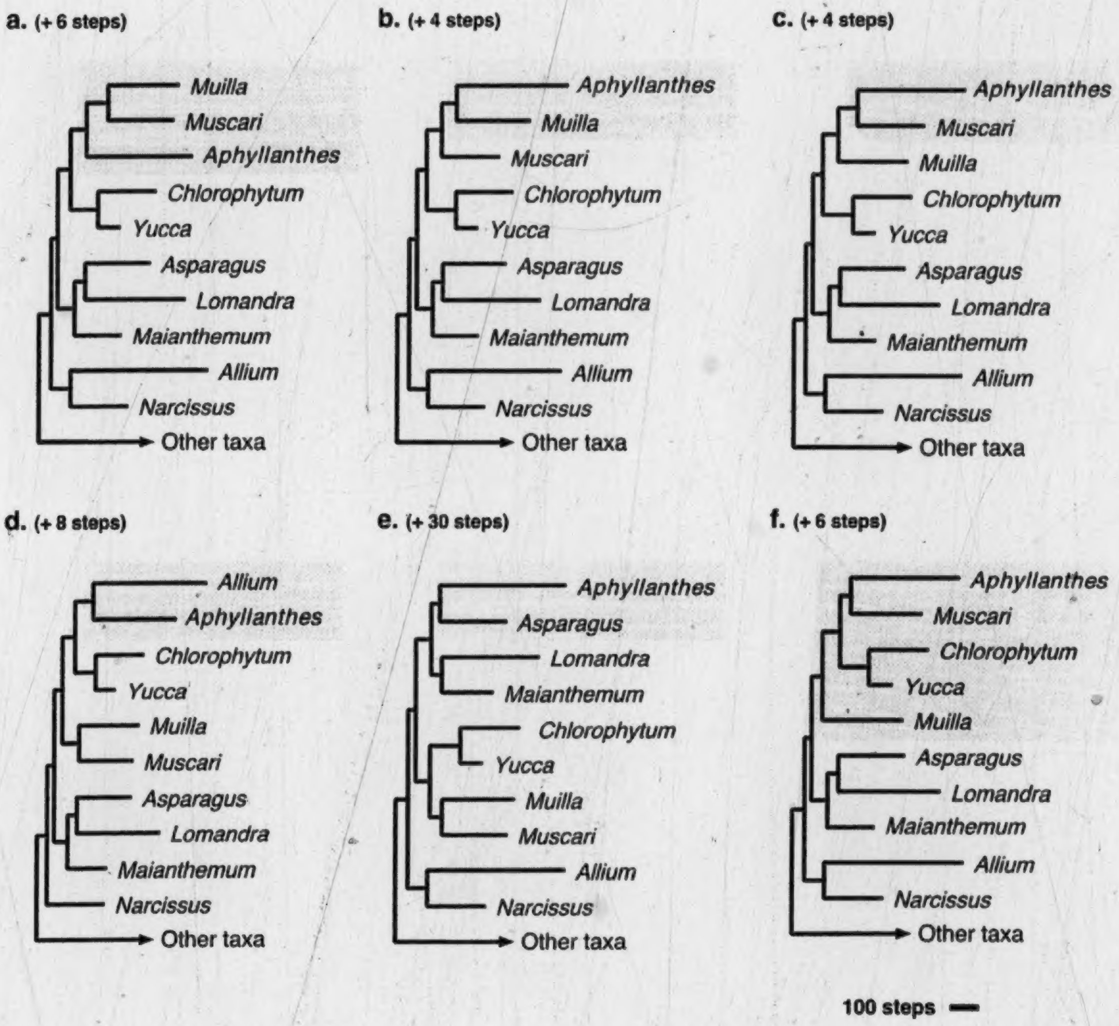

a. (+ 6 steps)
Muilla
Muscari
Aphyllanthes
Chlorophytum
Yucca
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Other taxa
b. (+ 4 steps)
Aphyllanthes
Muilla
Muscari
Chlorophytum
Yucca
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Other taxa
c. (+ 4 steps)
Aphyllanthes
Muscari
Muilla
Chlorophytum
Yucca
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Other taxa
d. (+ 8 steps)
Allium
Aphyllanthes
Chlorophytum
Yucca
Muilla
Muscari
Asparagus
Lomandra
Maianthemum
Narcissus
Other taxa
e. (+ 30 steps)
Aphyllanthes
Asparagus
Lomandra
Maianthemum
Chlorophytum
Yucca
Muilla
Muscari
Allium
Narcissus
Other taxa
f. (+ 6 steps)
Aphyllanthes
Muscari
Chlorophytum
Yucca
Muilla
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Other taxa
100 steps

FIG. 3.3. Single best maximum likelihood tree (ML score = 74997.65) inferred for Asparagales and three commelinoid outgroup taxa, based on a large chloroplast data set (see text and Fig. 3.1). The parameters of the assumed likelihood model (GTR + Γ + I; Table 3.3) were all preassigned with the following estimated values (see text for details): Base frequencies -- A = 0.29607, C = 0.17097, G = 0.19756, T = 0.33541; Substitution rate matrix, **R**, (AC = 1.5356; AG = 4.0685; AT = 0.46927; CG = 0.96500; CT = 5.0043; GT = 1); Proportion of invariable sites = 0.52556; Γ-shape parameter, α, = 0.74787. Phylogenetic relationships Asparagales are identical to the MP tree (Fig. 3.1) if the lineages indicated (*Asphodelus* and *Xanthorrhoea*) are interchanged with each other.

Figure 3.3

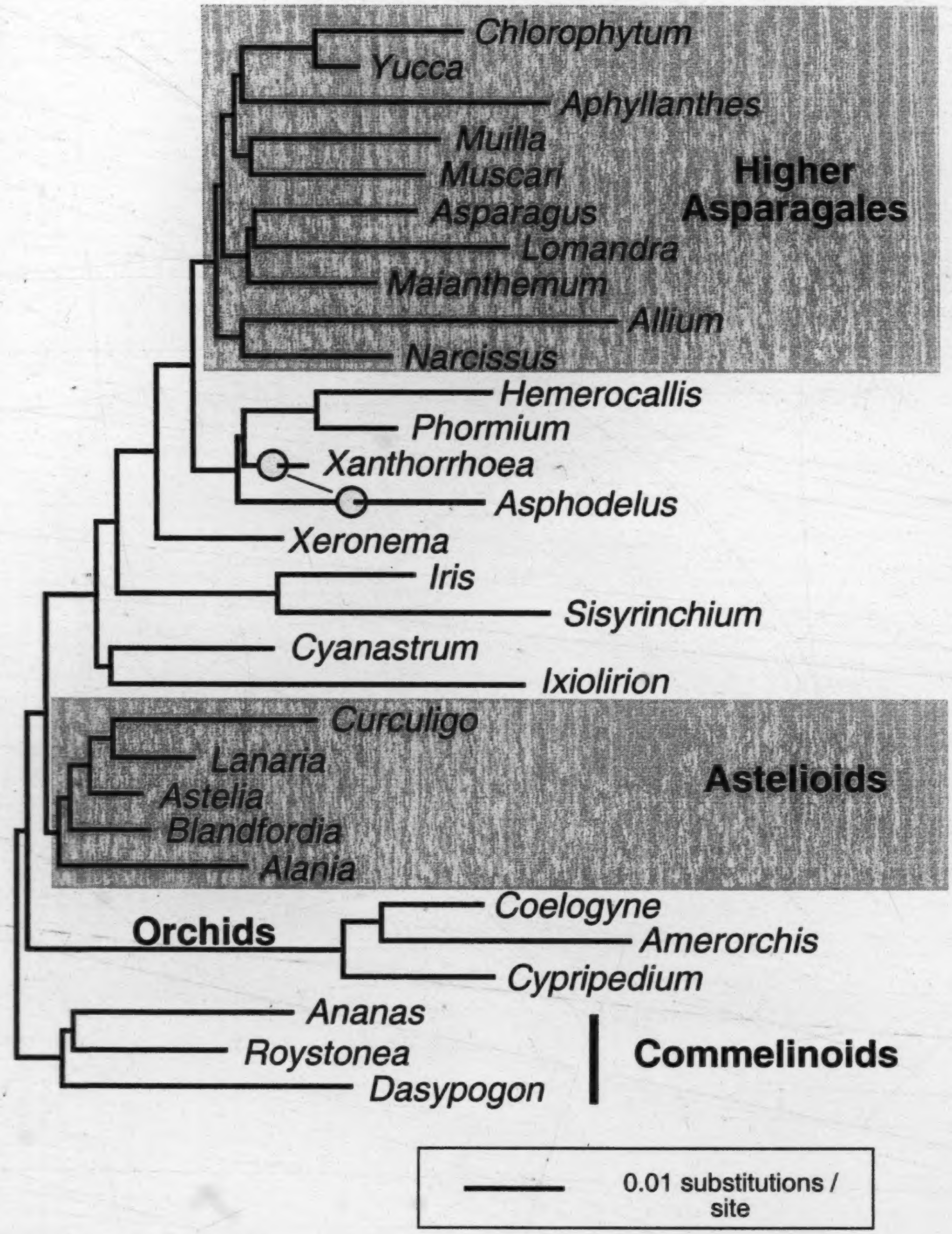

Chlorophytum
Yucca
Aphyllanthes
Muilla
Muscari
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Higher Asparagales
Hemerocallis
Phormium
Xanthorrhoea
Asphodelus
Xeronema
Iris
Sisyrinchium
Cyanastrum
Ixiolirion
Curculigo
Lanaria
Astelia
Blandfordia
Alania
Astelioids
Orchids
Coelogyne
Amerorchis
Cypripedium
Ananas
Roystonea
Dasypogon
Commelinoids
0.01 substitutions / site

FIG. 3.4. Branch labels scored in bootstrap analyses presented in Fig. 3.5 (topology of subtree as in Fig. 3.1).

Figure 3.4

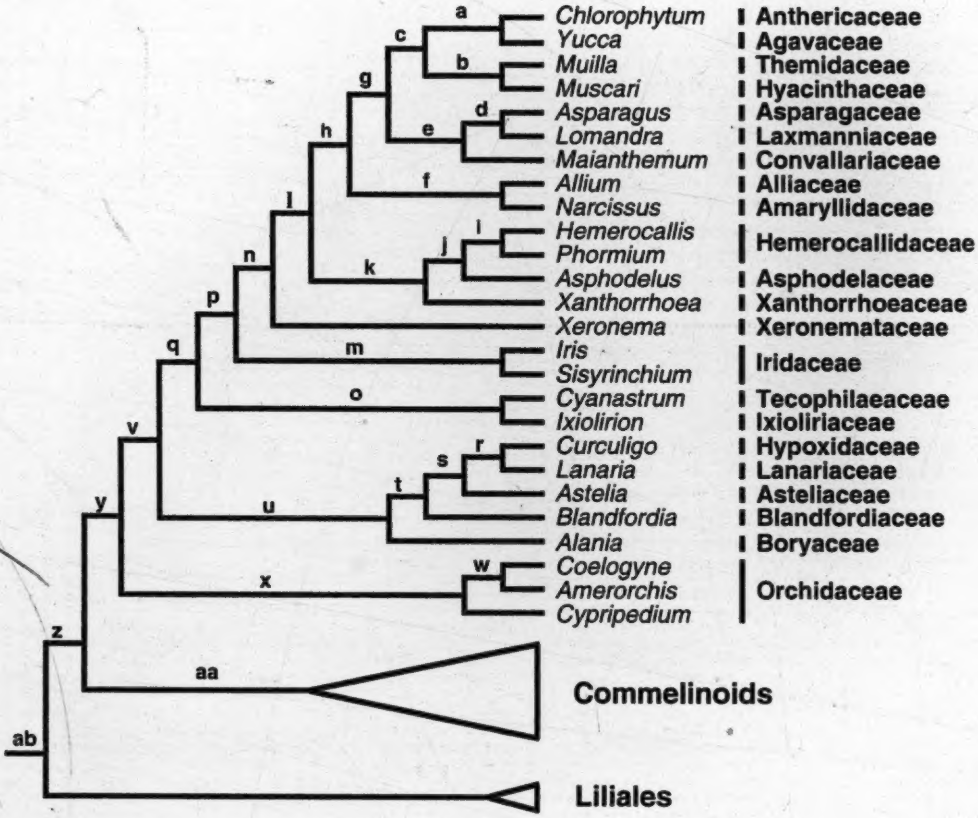
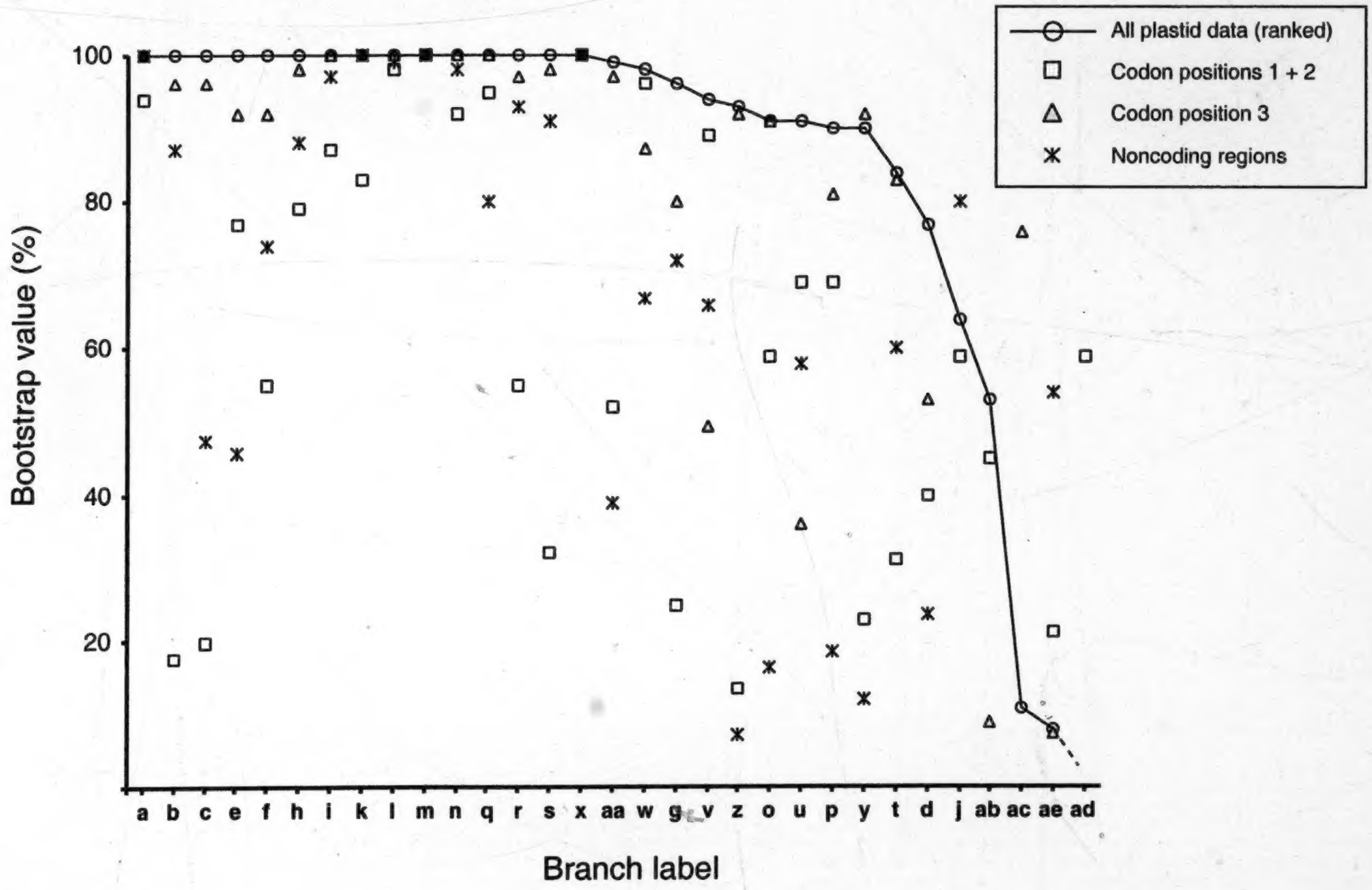

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
aa
ab
Chlorophytum
Yucca
Muilla
Muscari
Asparagus
Lomandra
Maianthemum
Allium
Narcissus
Hemerocallis
Phormium
Asphodelus
Xanthorrhoea
Xeronema
Iris
Sisyrinchium
Cyanastrum
Ixiolirion
Curculigo
Lanaria
Astelia
Blandfordia
Alania
Coelogyne
Amerorchis
Cypripedium
Anthericaceae
Agavaceae
Themidaceae
Hyacinthaceae
Asparagaceae
Laxmanniaceae
Convallariaceae
Alliaceae
Amaryllidaceae
Hemerocallidaceae
Asphodelaceae
Xanthorrhoeaceae
Xeronemataceae
Iridaceae
Tecophilaeaceae
Ixioliriaceae
Hypoxidaceae
Lanariaceae
Asteliaceae
Blandfordiaceae
Boryaceae
Orchidaceae
Commelinoids
Liliales

FIG. 3.5. Spectrum of bootstrap support for branches found in the most parsimonious tree (branch labels "a" through "ab"; Figure 3.4), and a few additional branches that had > 50% support from at least one data partition ["ac" = (*Asphodelus* + *Xanthorrhoea*); "ad" = (orchids + Liliiales); "ae" = (*Ixiolirion* + Iridaceae)]. Four bootstrap analyses are summarized here: all of the plastid data combined, codon positions 1 and 2 combined, codon position 3 combined, and all non-coding regions combined. Branches are ranked according to their support from all of the data combined (i.e., codon positions 1-3 plus all other coding plus noncoding regions). Bootstrap values < 5% are ignored.

Figure 3.5



All plastid data (ranked)
Codon positions 1 + 2
Codon position 3
Noncoding regions
Bootstrap value (%)
100
80
60
40
20
a b c e f h i k l m n q r s x aa w g v z o u p y t d j ab ac ae ad
Branch label

Bibliography

Angiosperm Phylogeny Group., 1998. An ordinal classification for the families of flowering plants. Ann. Missouri Bot. Gard. 85, 531–553.

Barker, F. B., and Lutzoni, F. M., 2002. The utility of the incongruence length difference test. Syst. Biol. 51, 625–637.

Barry, D., and Hartigan, J. A., 1987. Asynchronous distance between homologous DNA sequences. Biometrics 43, 261–276.

Chang, J. T., 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. Math. Biosci. 134, 189–215.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedrén, M., Gaut, B. S., Jansen, R. K., Kim, K.-J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q.–Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, Jr. G. H., Graham, S. W., Barrett, S. C. H., Dayanandan, S., and Albert, V. A,, 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80, 528–580.

Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J.G., Cox, A. V., Eguiarte, L. E., Hartwell, J., Fay, M. F., Caddick, L. R., Cameron, K. M., and Hoot, S., 1995. Molecular phylogenetics of Lilianae. In: Rudall, P. J., Cribb, P. J., Cutler, D. F., Humphries, C. J., (Eds.), Monocotyledons: Systematics and Evolution. Royal Botanic Gardens, Kew, pp. 109–137.

Chase, M. W., Rudall, P. J., and Conran, J. G., 1996. New circumscriptions and a new family of asparagoid lilies: genera formerly included in Anthericaceae. Kew Bull. 51, 667–680.

Chase, M. W., Rudall, P. J., and Conran, J. G., 1997. Validation of the family name Boryaceae. Kew Bull. 52, 416.

Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., Sullivan, S., Joseph, J., Molvray, M., Kores, P.J., Givnish, T. J., Sytsma, K. J., and Pires, J. C., 2000a. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 3–16.

Chase, M. W., Rudall, P. J., Fay, M. F., and Stobart, K. L., 2000b. Xeronemataceae, a new family of asparagoid lilies from New Caledonia and New Zealand. Kew Bull. 55, 865–870.

Chiang, T.-Y., Schaal, B. A., and Peng, C. I., 1998. Universal primers for amplification and sequencing a noncoding spacer between the *atp*B and *rbc*L genes of chloroplast DNA. Bot. Bull. Acad. Sin. (Taipei). 39, 245–250.

Conran, J. G., Chase, M. W., and Rudall, P. J., 1997. Two new monocotyledon families: Anemarrhenaceae and Behniaceae (Lilianae: Asparagales). Kew Bull. 52, 995–999.

Conran, J. G., 1998. Anthericaceae. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp. 114–121.

Cronquist, A., 1981. An integrated system of classification of flowering plants. Columbia University Press. N.Y. USA.

Cronquist, A., 1988. The evolution and classification of flowering plants, second edition. The New York Botanical Garden, Bronx, N.Y. USA. pp. 451–517.

Dahlgren, R. M. T., 1980. A revised system of classification of the angiosperms. Bot. J. Linn. Soc. 80, 91–124.

Dahlgren, R. M. T., and Clifford, H. T., 1982. The monocotyledons: a comparative study. Springer-Verlag, Berlin.

Dahlgren, R. M. T., and Rasmussen, F. N., 1983. Monocotyledon evolution: characters and phylogenetic estimation. Evol. Biol. 16, 255–395.

Dahlgren, R. M. T., Clifford, H. T., and Yeo, P. F., 1985. The families of monocotyledons: structure, evolution, and taxonomy. Springer-Verlag, Berlin.

Donoghue, M. J., 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43, 1137–1156.

Doyle, J. J., and Doyle, J. L., 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Duvall, M. R., Learn, Jr., G. H., Eguiarte, L. E., and Clegg, M. T., 1993. Phylogenetic analysis of *rbc*L sequences identifies *Acorus calamus* as the primal extant monocotyledon. Proc. Natl. Acad. Sci. USA. 90, 4641–4644.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C., 1994. Testing significance of incongruence. Cladistics 10, 315–319.

Fay, M. F., and Chase, M. W., 1996. Resurrection of Themidaceae for the *Brodiaea* alliance, and recircumscription of Alliaceae, Amaryllidaceae and Agapanthoideae. Taxon 45, 441–451.

Fay, M. F., Rudall, P. J., Sullivan, S., Stobart, K. L., de Bruijn, A. Y., Reeves, G., Qamaruz-Zaman, F., Hong, W. -P., Joseph, J., Hahn, W. J., Conran, J. G., and Chase, M. W., 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 360–371.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368-376.

Felsenstein, J., 1983. Parsimony in systematics: biological and statistical issues. Ann. Rev. Ecol. Syst. 14, 313–333.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783–791.

Graham, S. W., Kohn, J. R., Morton, B. R., Eckenwalder, J. E., and Barrett, S. C. H., 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. Syst. Biol. 47, 545–567.

Graham, S. W., and Olmstead, R. G., 2000a. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Amer. J. Bot. 87, 1712–1730.

Graham, S. W., and Olmstead, R. G., 2000b. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. Curr. Genet., 37, 183–188.

Graham, S. W., Reeves, P. A., Burns, A. C. E., Olmstead, R. G., 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int. J. Plant Sci. 161, S83–S96.

Hasegawa, M., Kishino, H., Yano T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21, 160–174.

Hendy, M. D., and Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Hillis, D. M., and Bull, J. J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Huber, H., 1969. Die Samenmerkmale und Verwandtschaftsverhältnisse der Liliiflorae. Mitteilungen der botanischen Staatssammlung München 8, 219–538.

Huelsenbeck, J. P., and Crandall, K. A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Ann. Rev. Ecol. Syst. 28, 437–466.

Hoot, S. B., Culham, A., and Crane, P. R., 1995. The utility of *atp*B gene sequences in resolving phylogenetic relationships: comparison with *rbc*L and 18S ribosomal DNA sequences in Lardizabalaceae. Ann. Missouri Bot. Gard. 82, 194–207.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., and Donoghue, M. J., 2002. Plant systematics: a phylogenetic approach, Second edition. Sinauer Associates, Sunderland, MA.

Jukes, T. H., and Cantor, C. R., 1969. Evolution of protein molecules. In: Munro, H. N. (Ed.), Mammalian Protein Metabolism. Academic Press, New York, pp. 21-132

Kim, K.-J., and Jansen, R. K., 1995. *ndh*F sequence evolution and the major clades in the sunflower family. Proc. Natl. Acad. Sci. USA 92, 10379–10383.

Kite, G. C., Grayer, R. J., Rudall, P. J., and Simmonds, M. S. J., 2000. The potential for chemical characters in monocotyledon systematics. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 101–113.

Krause, K., 1930. Liliaceae. In Engler, A., (Ed.), Die natürlichen Pflanzenfamilien, Second edition, Volume 15a. W. Engelmann, Leipzig, pp. 227–386.

Kubitzki, K., 1998a. The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York.

Kubitzki, K., 1998b. Agapanthaceae. In: Kubitzki, K., (Ed.), The Families and Genera of Vascular Plants, Volume III, Flowering plants. Monocotyledons: Lilianae (except Orchidaceae). Springer-Verlag, New York, pp. 58–60.

Lanave, C., Preparata, G., Saccone, C., and Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

Naylor, G. J. P., and Brown, W. M., 1998. *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47, 61–76.

Neyland, R., and Urbatsch, L. E., 1996a. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndh*F chloroplast gene sequences. Amer. J. Bot. 83, 1195–1206.

Neyland, R., and Urbatsch, L. E., 1996b. The *ndh*F chloroplast gene detected in all vascular plant divisions. Planta 200, 273–277.

Olmstead, R. G., and Sweere, J. A., 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43, 467–481.

Olmstead, R. G., and Reeves, P. A., 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. Ann. Missouri Bot. Gard. 82, 149–175.

Pires, J. C., Fay, M. F., Davis, W. S., Hufford, L., Rova, J., Chase, M. W., and Sytsma, K. J., 2001. Molecular and morphological phylogenetic analyses of Themidaceae (Asparagales). Kew Bull. 56, 601–626.

Pires, J. C., and Sytsma, K. J., 2002. A phylogenetic evaluation of a biosystematic framework: *Brodiaea* and related petaloid monocots (Themidaceae). Amer. J. Bot. 89, 1342–1359.

Rambaut, A., 1998. "Se-Al (Sequence Alignment Editor Version 1.0)," Computer program and documentation. Department of Zoology, University of Oxford, UK.

Rambaut, A., and Grassly, N. C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235–238.

Rodríguez, F., Oliver, J. L., Marín, A., and Medina, J. R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Rudall, P. J., Furness, C. A., Chase, M. W., Fay, M. F., 1997. Microsporogenesis and pollen sulcus type in Asparagales (Lilianae). Can. J. Bot. 75, 408–430.

Rudall, P. J., Chase, M. W., Cutler, D. F., Rusby, J., and De Bruijn, A. Y., 1998. Anatomical and molecular systematics of Asteliaceae and Hypoxidaceae. Bot J. Linn. Soc. 127, 1–42.

Sanderson, M. J., Wojciechowski, M. F., Hu, J.-M., Sher Khan, T., and Brady, S. G., 2000. Error, bias and long branch attraction in data for two chloroplast genes in seed plants. Mol. Biol. Evol. 17, 782–797.

Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., Bayer, C., Fay, M. W., de Bruijn, A. Y., Sullivan, S., and Qiu, Y., 2000. Phylogenetics of flowering plants based upon a combined analysis of plastid *atp*B and *rbc*L gene sequences. Syst. Biol. 49, 306–362.

Savolainen, V., Chase, M. W., Salamin, N., Soltis, D. E., Soltis, P. S., López, A. J., Fédrigo, O., and Naylor, G. J. P., 2002. Phylogeny reconstruction and functional constraints in organellar genomes: plastid *atp*B and *rbc*L sequences versus animal mitochondrion. Syst. Biol. 51, 638–647.

Swofford, D. L., Waddell, P. J., Huelsenbeck, J. P., Foster, P. G., Lewis, P. O., and Rogers, J. S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.

Swofford, D. L., 2002. "PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4," Computer program and documentation. Sinauer Associates, Sunderland, MA.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lec. Math. Life Sci. 17, 57–86.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M., 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA. 91, 9794–9798.

Zurawski, G., Clegg, M. T., and Brown, A. H. D., 1984. The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106, 735–749.

~~ CHAPTER FOUR ~~

CONCLUSIONS AND FUTURE DIRECTIONS

Significance of the results.

The work described here provides a very robust phylogenetic framework for the monocot order Asparagales, based on a large plastid data set sampled for exemplar species that represent its major evolutionary lineages. It largely upholds recent molecular work on the order and its relatives by Chase *et al.* (2000), Fay *et al.* (2000) and others, and broadly verifies the circumscription of the order in the morphology-based classification scheme of Dahlgren *et al.* (1985). An opposing morphology-based classification scheme is rejected, that of Cronquist (1981, 1988), who lumped many of these taxa into a single taxon, the family Liliaceae (Liliales).

In Chapter 3, I provide the first clear evidence for a sister-group relationship between Asparagales and a large cluster of monocot orders associated with the grasses, palms, gingers and relatives, the so-called commelinoid monocots (*sensu* APG, 1998), and also for the precise branching order at the base of Asparagales. My data also firmly support the inclusion of two large families in the order, Orchidaceae and Iridaceae, that were excluded by Dahlgren *et al.* (1985) but included in Asparagales in recent molecular studies. The orchids emerge at the basal split in the order, followed by a cluster of five small, primarily Southern Hemisphere families, the "astelioids" (Chase *et al.*, 1996; Rudall *et al.*, 1998; Fay *et al.* 2000), whose inter-relationships are firmly resolved by my data. Relationships among a large cluster of families comprising the "higher asparagoids" (Chase *et al.*, 1995; Rudall *et al.*, 1997) are also clarified and firmly resolved here, with one exception, concerning the monotypic family Aphyllanthaceae (*Aphyllanthes monspeliensis* L.).

When *Aphyllanthes* L. is included in phylogenetic analysis it disrupts phylogenetic inference in the higher Asparagales, as noted by Fay *et al.* (2000). The destabilizing effect of this taxon on phylogenetic inference was explored with several methods, including Monte Carlo simulation. I showed that if the observed placement of *Aphyllanthes* (as the sister group of Agavaceae-Anthericaceae; Chapter 3, Fig. 3.1) is correct, it would be strongly inferred. However, my data are also shown to be prone to inferring this relationship if another evolutionary scenario (with *Aphyllanthes* as the sister group of Hyacinthaceae-Themidaceae) is actually the correct topology. Thus, I cannot chose between these hypotheses using maximum parsimony analysis of the current data.

Despite uncertainty over the position of *Aphyllanthes*, the large majority of branches on the backbone of Asparagales phylogeny are resolved with very substantial bootstrap support (an average of 95% bootstrap support across the branches examined here). I did several analyses of phylogenetic congruence among different subsets of the plastid data, breaking these down into different codon positions, or coding versus noncoding DNA. These analyses demonstrate remarkable agreement among the different data subsets, with no evidence of substantially incongruent signals among them.

One of the few relationships to be poorly supported concerns a clade of three closely related families, Asphodelaceae, Hemerocallidaceae and Xanthorrhoeaceae. Various analyses in Chapter 2 and 3 yield the three possible arrangement of these taxa. A survey summarized in Chapter 2 reveals that two of these families (Asphodelaceae and Hemerocallidaceae) have taxa lacking an intron in a slowly evolving plastid gene, 3'-*rps*12. The cost of different loss scenarios is explored in detail in Chapter 2, but, rather surprisingly, I conclude that the intron was likely lost in parallel, as it is missing in all members of Asphodelaceae examined, but only a subset of Hemerocallidaceae. I show that intron absence is not a marker of monophyly for all of the taxa in Asphodelaceae and Hemerocallidaceae that lack it. However, it adds further support to the synonymy of another family, Johnsoniaceae, with Hemerocallidaceae, because all taxa examined for the former family lack the intron, as do a subset of Hemerocallidaceae (*Hemerocallis* and *Simethis*).

Directions.

My study provides a robust historical framework for evolutionary studies in Asparagales, both for hypothesis testing (see Donoghue, 1989, for example) and for focussing searches for morphological characters (synapomorphies) that may support major branches in the order (see Judd *et al.*, 2002, for examples). Future work on Asparagales higher-order systematics using the regions I examined should also focus on including lineages not sampled here. The deepest missing lineage is probably Doryanthaceae, but other families that likely reside in the higher Asparagales also should be included (Agapanthaceae, Anemarrhenaceae, Behniaceae, Herreriaceae and Hesperocallidaceae). It would probably also be profitable to include more representatives of some of the larger families, both to clarify intrafamilial relationships, and to help break up some of the longer deep branches.

It would also be interesting to pursue the molecular evolution of the truncated gene (*rps*7 in *Lomandra longifolium*) and pseudogenes (*ndh*F and *ndh*B in some orchids) encountered in the course of the study, to determine how phylogenetically widespread they are, and to explore their evolutionary degeneration or alteration. In addition, it would be of interest to determine whether the genes for the other subunits of plastid NADH dehydrogenase have also been lost.

One analytical puzzle that requires further attention concerns the position of Aphyllanthaceae in the higher Asparagales. Fay *et al.* (2000) suggested that the mystery of its placement would only be resolved with additional data. However, if, as seems likely, its distorting effects on phylogeny inference are a consequence of it being a problematic "long-branch" taxon, then the addition of new data may ultimately lead to a very robust but wrong result, the problem of "statistical inconsistency" (Felsenstein, 1978, Hendy and Penny, 1985). Instead it may be necessary to analyze the data in a smarter way. For example, the approach I used in Chapter 3, Monte Carlo simulation (following Sanderson *et al.*, 2000), could potentially be extended to determine whether maximum likelihood analysis is less prone to yielding biased results concerning the position of *Aphyllanthes* than maximum parsimony.

Bibliography

Angiosperm Phylogeny Group, 1998. An ordinal classification for the families of flowering plants. Ann. Missouri Bot. Gard. 85, 531–553.

Chase, M. W., Rudall, P. J., and Conran, J. G., 1996. New circumscriptions and a new family of asparagoid lilies: genera formerly included in Anthericaceae. Kew Bull. 51, 667–680.

Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J.G., Cox, A. V., Eguiarte, L. E., Hartwell, J., Fay, M. F., Caddick, L. R., Cameron, K. M., and Hoot, S., 1995. Molecular phylogenetics of Lilianae. In: Rudall, P. J., Cribb, P. J., Cutler, D. F., Humphries, C. J., (Eds.), Monocotyledons: Systematics and Evolution. Royal Botanic Gardens, Kew, pp. 109–137.

Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., Sullivan, S., Joseph, J., Molvray, M., Kores, P.J., Givnish, T. J., Sytsma, K. J., Pires, J. C., 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 3–16.

Cronquist, A., 1981. An integrated system of classification of flowering plants. Columbia University Press. N.Y. USA.

Cronquist, A., 1988. The evolution and classification of flowering plants, Second edition. The New York Botanical Garden, Bronx, N.Y., USA. pp. 451–517.

Dahlgren, R. M. T., Clifford, H. T., and Yeo, P. F., 1985. The families of monocotyledons: structure, evolution, and taxonomy. Springer-Verlag, Berlin.

Donoghue, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43: 1127-1156.

Fay, M. F., Rudall, P. J., Sullivan, S., Stobart, K. L., de Bruijn, A. Y., Reeves, G., Qamaruz-Zaman, F., Hong, W. -P., Joseph, J., Hahn, W. J., Conran, J. G., and Chase, M. W., 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. In: Wilson, K. L., Morrison, D. A., (Eds.), Monocots: Systematics and Evolution. CSIRO, Melbourne, pp. 360–371.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Hendy, M. D., and Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., Donoghue, M. J., 2002. Plant systematics: a phylogenetic approach, Second edition. Sinauer Associates, Sunderland, MA.

Rudall, P. J., Furness, C. A., Chase, M. W., and Fay, M. F., 1997. Microsporogenesis and pollen sulcus type in Asparagales (Lilianae). Can. J. Bot. 75, 408–430.

Rudall, P. J., Chase, M. W., Cutler, D. F., Rusby, J., and De Bruijn, A. Y., 1998. Anatomical and molecular systematics of Asteliaceae and Hypoxidaceae. Bot J. Linn. Soc. 127, 1–42.

Sanderson, M. J., Wojciechowski, M. F., Hu, J.-M., Sher Khan, T., and Brady, S. G., 2000. Error, bias and long branch attraction in data for two chloroplast genes in seed plants. Mol. Biol. Evol. 17, 782–797.